Editor: Dana Simian

Proceedings of The International Conference on Applied Informatics

ICDD

IMAGINATION, CREATIVITY, DESIGN, DEVELOPMENT

Romania, Sibiu, May 16 - 18, 2019

IMAGINATION, CREATIVITY, DESIGN, DEVELOPMENT

Proceedings of the International Conference on Applied Informatics
ICDD

May 16$^{\text{th}}$ – 18$^{\text{th}}$, 2019
Sibiu, Romania

Lucian Blaga University of Sibiu, 2019

Editor Dana Simian

All papers in this volume were peer review by two independent reviewers

Associated Editor   Laura Florentina Stoica

Proceedings of the International Conference on Applied Informatics, ICDD
May 16th – 18th, 2019, Sibiu, Romania

Preface

This volume contains refereed papers presented within the International Conference on Applied Informatics, "Imagination, Creativity, Design, Development" - ICDD 2019, which was held between May 16th – 18th, at the Faculty of Sciences, Lucian Blaga University of Sibiu, Romania.

The conference is mainly addressed to young researchers from all over the world. The conference gives the participants the opportunity to discuss and present their research on informatics and related fields (like computational algebra, numerical calculus, bioinformatics, etc.). The conference welcomes submissions of original papers on all aspects of informatics and related fields ranging from new concepts and theoretical developments to advanced technologies and innovative applications. Specific topics of the conference included but were not restricted to: Algorithms and Data Structures, Graph Theory and Applications, Formal Languages and Compilers, Cryptography, Modelling and Simulation, Computer Programming, Computer Vision, Computer Graphics, Game Design, Data Mining, Distributed Computing, Artificial Intelligence, Service Oriented Applications, Networking, Grid Computing, Mobile Operating Systems, Scientific Computing, Software Engineering, Bioinformatics, Robotics, Computer Architecture, Evolutionary Computing, Multimedia Systems, Internet Communication and Technologies, Web Applications.

The conference has brought together participants from 8 countries (Bulgaria, Germany, Italy, Romania, Russia, Serbia, Ukraine and United States of America).
We thank all the participants for their interesting talks and discussions. We also thank the members of the scientific committee for their help in reviewing the submitted papers and for their contributions to the scientific success of the conference.

December 2019   Dana Simian
                                                        Conference chairman

Motto:

"There are no limits, only your imagination"

Scientific committee

Kiril Alexiev - Bulgarian Academy of Sciences, Bulgaria
Vsevolod Arnaut - Moldova State University, Republic of Moldova
Alina Barbulescu - Higher College of Technology, Sharjah, UAE
Lasse Berntzen - Buskerud and Vestfold University College, Norway
Peter Braun - University of Applied Sciences, Würzburg-Schweinfurt, Germany
Uvgrkcp'Ekwtgc"/'Nwekcp'Dnci c'Wpkxgtukv{'qh'Ukdkw.'Tqo cpkc"
Nicolae Constantinescu - University of Craiova, Romania
Daniela Danciulescu - University of Craiova, Romania
Oleksandr Dorokhov - Kharkiv National University of Economics, Ukraine
Ralf Fabian - Lucian Blaga University of Sibiu, Romania
Stefka Fidanova - Bulgarian Academy of Sciences, Bulgaria
Ulrich Fiedler - Bern University of Applied Science, Switzerland
Adrian Florea - Lucian Blaga University of Sibiu, Romania
Teresa Gonçalves - University of Evora, Portugal
Andrina Granić - University of Split, Croatia
Katalina Grigorova - University of Ruse, Bulgaria
Daniel Hunyadi - Lucian Blaga University of Sibiu, Romania
Saleema JS - Chris University, Bangalore, India
Milena Lazarova - Technical University of Sofia, Bulgaria
Lixin Liang - Tsinghua University, Beijing, China
Suzana Loskovska - Ss. Cyril and Methodius University, Republic of Macedonia in Skopje
Rossitza S. Marinova - Concordia University of Edmonton, Canada
Gabriela Moise - Petroleum-Gas University of Ploiesti, Romania
G.Jose Moses - Raghu Engineering College Visakhapatnam, Andhra Pradesh, India
Mircea Musan - Lucian Blaga University of Sibiu, Romania
Mircea Iosif Neamtu - Lucian Blaga University of Sibiu, Romania
Grażyna Paliwoda-Pękosz - Cracow University of Economics, Poland
Camelia Pintea - North University Center of Baia Mare, Romania
Antoniu Pitic - Lucian Blaga University of Sibiu, Romania
Alina Pitic - Lucian Blaga University of Sibiu, Romania
Anca Ralescu - University of Cincinnati, United States of America
Mohammad Rezai - Sheffield Hallam University, United Kingdom
José Saias - University of Evora, Portugal
Livia Sangeorzan - Transilvania University of Brasov, Romania
Klaus Bruno Schebesch - Vasile Goldis University, Arad, Romania
Soraya Sedkaoui - Khemis Miliana University, Algeria
Dana Simian - Lucian Blaga University of Sibiu, Romania
Ansgar Steland - RWTH Aachen University, Germany
Florin Stoica - Lucian Blaga University of Sibiu, Romania
Laura Florentina Stoica - Lucian Blaga University of Sibiu, Romania

Grażyna Suchacka - University of Opole, Poland

Jolanta Tańcula - University of Opole, Poland

Milan Tuba - Singidunum University, Belgrade, Serbia

Anca Vasilescu - Transilvania University of Brasov, Romania

Dana Vasiloaica - Institute of Technology Sligo, Ireland

Sofia Visa - The College of Wooster, United States of America

Contents

Bulk email sender – the ultimate spammer experience?

Vasile'Cpi j gnw , Eduard-Traian Ștefănescu, Felix Husac

Abstract

The aim of this article is to propose a solution for the improvement of the current way of sending e-commerce related emails and possibly help declutter the end-user's inbox. A problem of interest in all companies and providers of services and products is to keep a large group of people up to date with their products or services. We propose a bulk email sender, so that businesses can find it easier to send out sales ads, special promos or promo codes to selected customers or just keep everyone up to date. Our application has a simple, clean and very user-friendly interface.

1   Introduction

This article is the result of the need for a fast, easy-to-use and reliable app to manage email communications within colleges, universities and even e-commerce related businesses. The current way of creating email groups or even using different bulk email sender applications is the cause of many headaches and mishaps. By offering a new, simple and most importantly easy to use bulk email sender, many problems that users face (the need to copy address lists, create hundreds of mailing groups or phone people to confirm that they received a certain email) are significantly reduced.

At present, there are many implementations of this idea, both paid and for free. For example, other options could be "e-Campaign" (but it has a price tag of 147 USD and the trail version is rather limiting), "Atomic Mail Sender" (has limitations on the trial version and costs 80 USD) or "SendBlaster" (again, the trial has some key functions missing versus the paid app), just to name a few.

One of the benefits of e-Campaign is their long presence on the market, meaning maintenance and updates for their software. The major problem is that, in their 19+ years of activity, while the technical, back-end side of their bulk email sender grew and developed into a complex, complete system, the design of their app has remained almost the same. Although it encompasses the main roles such an app should play, and many more, most first time users find it rather intimidating for a beginner. Also, the free version of the app only lets you send a maximum of 50 emails per job and has a watermarking feature, appending a footer after every email, while all more advanced functions are reserved for the paid app (costing the hefty sum of 149.95 EUR for Standard Edition and 279.95 EUR for the Business Edition).

This is a reoccurring theme amongst the products listed above. Another example, the Atomic Mail Sender bulk email sender, although offering fewer functions than e-Campaign, costs about the same. It has a slightly friendlier user interface, with better organised buttons and menus, but it too can be quite overwhelming for a first time user.

Our last choice, and the best looking of all three, is SendBlaster. It is really well-organised, offers a wide range of functions (although not as many as e-Campaign) but , because of it's price tag (106 USD), we felt we can do better. After all, not everybody needs 20 bonus HTML templates, Google Analytics integration and so forth.

In a nutshell, most current solutions are difficult to use, have an ancient looking user interface or are just too expensive for what they do.

We decided to only focus around the most crucial aspects a bulk email sender should have and ignoring what didn't seem useful for our specific use cases.

The rest of the article is organized as follows. In Section 2, we describe the application design, treating both the GUI and the business logic behind the chosen design, and also the components that make up the application and how they work together, the future development plans. In Section 3, the conclusions reached while pursuing this project.

2 Application design

2.1 User interface design

In order to make a product that is free, reliable and easy to use, we propose in this paper a solution that we consider to be centered around the most important functions a bulk email sender should have:

- a clean dashboard, in order to see what emails were sent, to whom, when, and how many of the scheduled emails have been sent since the timer was set
- a simple sender dashboard, similar to Gmail and yahoo Send interfaces, for on the fly composing, editing and sending emails
- a scheduler, for organizing and prioritizing email sendouts
- a way to import readymade email lists, such as microsoft Excell documents (.xlxs or .csv files)

To achieve the set goals, the development of the app was divided into three parts:

- the design of the user interface
- the client application (the app that the user interacts with)
- the server (the component that would take care of sending the emails)

2.1.1 Dashboard tab

The "Dashboard" tab is the main tab of the app, it allows the user to see when a certain email was sent, to whom, what the subject was and also, give the user a glimpse of the subject. At the bottom of the screen, a counter specifies the number of emails sent for a certain subject. If, from the displayed list, an email is selected, a red "X" will appear on its right-hand side. This allows the user to delete the email form the screen, regardless if it was already sent, or waiting to be sent to the recipient. The user can still find the email in the database, but those emails that were deleted before being sent are flagged as "cancelled". Fig. 1 illustrates the Dashboard tab, where the To, Subject and Content fields for the sent emails are visible. The Sent counter can also be seen at the bottom of Fig. 1.
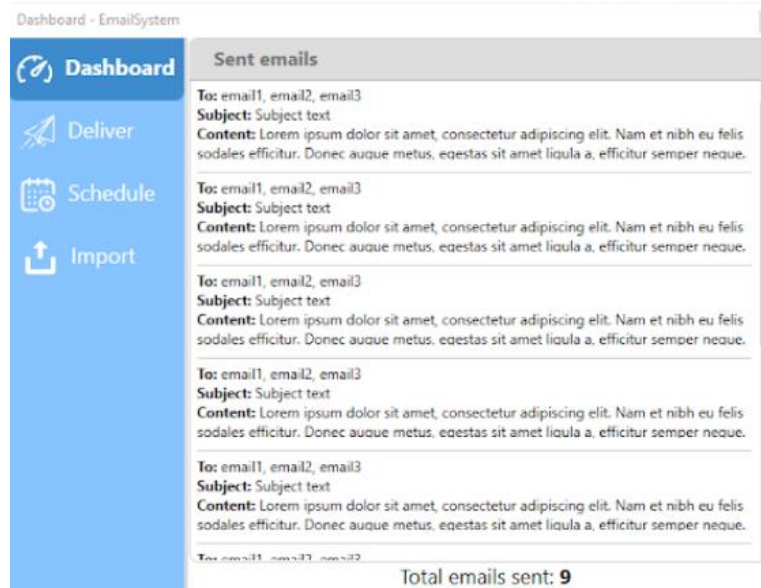
Fig. 1 Dashboard Tab

2.1.2 Deliver tab

The "Deliver" tab is just a simple text editing dashboard, similar in layout and function to other web based email clients, containing a "To", "Subject" and "Content" fields. Fig. 2 is a screenshot of the Deliver tab, showing the standard layout of an email composer.
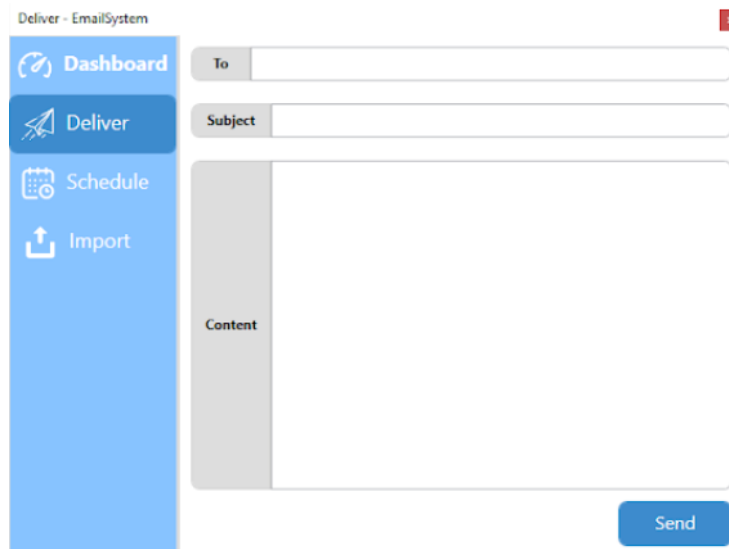
Fig. 2 Deliver Tab

2.1.3 Schedule tab

The "Schedule" tab is a carbon-copy of the previous window, save for a scrollable list/textbox from where the user can pick a time and a date for the delivery, or they can type it in themselves. Multiple emails can be sent from this tab at the same time, just by altering the subject, content or recipients fields. The date and time field can be left alone, because the values placed into it do not change, unless explicitly modified (by the user). Fig. 3 presents the Schedule tab. The difference between this screen and the previous one, Deliver, consists of the "When" spinner/edit box, located at the bottom of the screen.
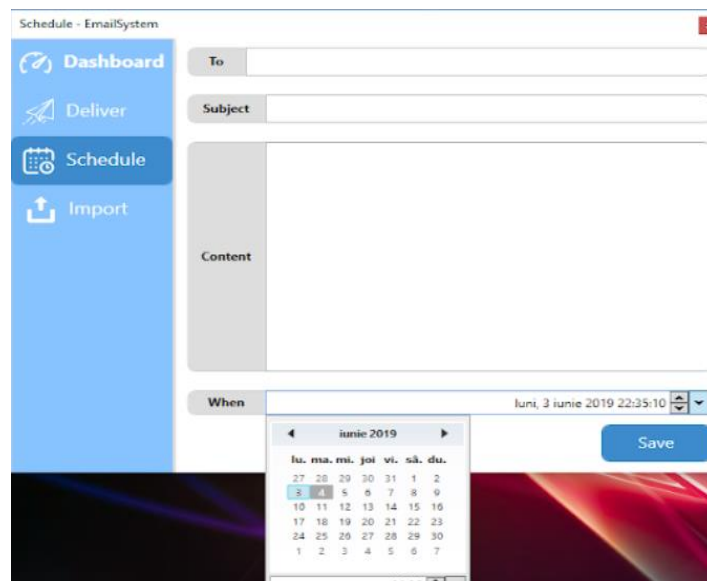
Fig. 3 Scedule Tab

2.1.4 Import tab

The "Import" tab is the one used for uploading ready made email templates into the app. We support both .xlsx and .csv files, from which the app "reads" the subjects, contents and recipients of as many emails as the user wishes and then, by pressing "Send", starts sending them one by one. Fig. 4 shows the file chooser panel that pops up when selecting "Import" from the left-hand side app menu. The app screen remains blank until the user picks the file they need. After the selection was made, the emails are loaded on the page, waiting to be sent or scheduled.
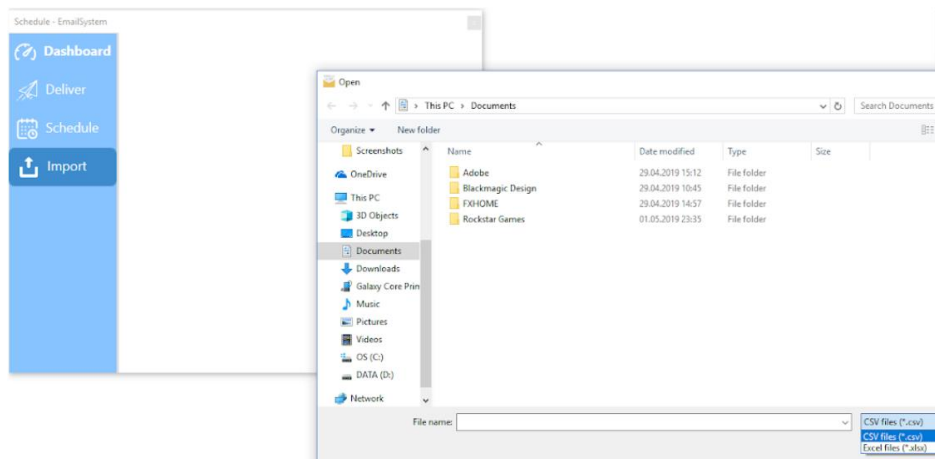
Fig. 4 Import Tab

The email templates are easy to generate. For example, such an email template can be made by creating a spreadsheet in Microsoft Excell, Google Spreadsheets etc. and naming the first three columns (A1-A3) "To", "Title" and "Subject".

Users can add as many email addresses as they wish, separated by a comma, in the row (or rows) below the "To" table heading. Adding addresses in the same cell means that everyone added there will get the same email (subject and content that are written in the cells next to the address one). Adding the addresses in different rows means that the recipients will receive different emails

(subject and content), so by just creating this simple way of organizing and sending emails, the need for countless email groups is eliminated.

This method of organizing the components of an email seems the most logical and easy to use way of managing email communications. Multiple copies of the list can be created, and each one altered for a specific purpose, with ready-made subjects and contents, or left blank, to be filled in later on.

If the user wishes, they can keep just one spreadsheet, and only update the recipients and contents of their emails by deleting and inserting new addresses and changing the subject and content of their emails inside the spreadsheet.

2.2 Business Logic And Technologies Used

For designing the user interface, an XML mockup design was created using draw.io, a free, web based design platform with a lot of flexibility and functionality. This approach was chosen in order to integrate the design with the C# client app.

The client side of the application was created using the C# programming language, together with WPF (Windows Presentation Foundation) standard, in order to design the app with XAML (a declarative markup language). Also, the client follows the MVVM [4] design pattern (also known as Model-View-Binder), a design pattern proprietary to Microsoft, that is especially helpful when designing event-driven programming of user interfaces. The benefits of using a declarative style language is the fact that the code is human readable, making it easier to manage the code. The MVVM pattern was used to separate the design (view), data (model) and the performed actions (view-model), for better control.

The server side of the application was written in Java, using SpringBoot, a variation of the Spring Java Framework, because of the embedded Tomcat server. The server is the component that does the actual sending of the email, so it is important for it to be easily written and easily maintained.

The communication between the two "sides" of the application is possible due to RestAPI. RestAPI [5] was used for the way it handles public API over the internet and the fact that it provides better support for CRUD operations than SOAP(at least for our use case).

The documentation for the methods being called from the client's side were automatically generated by RAML in an HTML document. RAML [6] stands for RESTful API Modeling Language and was used to design and describe the API.

To be able to send emails, an email address was needed for the sender (the application). Google`s G-mail service was used, and also Google`s SMTP server.

In order for the application to manage the different requirements of sending an email (direct sending or scheduled sending), a database was used. The database was created using MySQL. A secondary employment for the database is so that the user doesn't "feel" the load on his machine and internet connection, or doesn`t have to wait while the app sends the emails, because such an operation could be quite resource-expensive. The internet connection could be bad, other tasks running on a lower-end processor would increase the waiting time or the RAM memory would be already busy with say, a browser with multiple tabs opened. Another reason for using a database is to track the sent emails.

In order to understand the roles that each of the app`s components play in sending an email, a brief description of the two main possible scenarios: direct (instant) send or scheduled send will be presented.

If the user selects the "Instant email" tab, composes their email and hits "Send", the email first goes to the database. There, a thread checks if there are any unsent emails, and if there are, are they cancelled or not? (there is also the option to cancel the delivery of an email in the "Dashboard" tab by clicking on a red "X" that appears over a selected email).

If the server "canSend()" (the method that checks for valid emails) the email, then the email leaves. Also, if one of the addresses is invalid, an error message is displayed on a toast, asking the user to check the mailing addresses.

The same thing happens for scheduled emails, the only difference being that a second thread is created, and looks to see if any emails still remain in the database that are past their send-by date.

By using a database and checking it at fixed, 5 second intervals, the data being manipulated by our app can be better controlled. This allows the user a short window of time to cancel sending if a mistake occurred, and also avoids mixing up emails (instant sends that are not sent or scheduled emails being sent instantly etc.).

2.3 Future developments

The main concern regarding future iterations of the app is security. Not only for the people using the service to send emails, but also the recipients that trust that the emails they receive are relevant and important. Spam emails should be blocked and the account of the user that tried to send them should be terminated without the right of re-enrollment.

This will be achieved in future developments by implementing authentication and authorization safety steps. By doing so, no one with ill intentions can use the app - or if they do, they can get permanently banned from it. A proposition that requires further pursuit could be the requirement of a valid ID or Driver's License in order to sign up for the service. Not many people are willing to give away such valuable information, and the current risks of sensitive data management and storage outweigh the benefit of scaring away scammers and spammers. Better solutions need to be found.

Another issue that could arise is spam false-positives (when a valid email is flagged as Spam). A protocol for dealing with such emails should also be put in place.

Other future developments may include adding support for sending attachments and files and implementing different email designs and templates, to make the recipient's experience visually pleasing.

3 Conclusions

The aim of this article is to propose a solution to improve the current way of sending college or e-commerce related emails and possibly help declutter the end-user's inbox. A problem of interest in all companies and providers of services and products is to keep a large group of people up to date with their products or services. Another concern is keeping email communications relevant.

Our goal is not to give spammers other tools to send their emails, but to help legitimate businesses grow and maintain their email correspondence with clients and also communicate with their employees.

References

[1] *e-Camapign 12,* https://www.lmhsoft.com/ecamp/ (retrieved 26.07.2019)

[2] *Atomic Mail Sender,* https://www.atompark.com/bulk-email-sender-the-best-email-service-2016/ (retrieved 26.07.2019)

[3] *SendBlaster,* https://www.sendblaster.com/desktop-email-marketing-software-features (retrieved 26.07.2019)

[4] Jeremy Likness, *Model-View-ViewModel Explained,* https://www.codeproject.com/articles/100175/model-view-viewmodel-mvvm-explained (retrieved 26.07.2019)

[5] John Mueller, *Understanding SOAP and REST Basics and Differences,* https://www.smartbear.com/blog/test-and-monitor/understanding-soap-and-rest-basics (retrieved 26.07.2019)

[6] Jitendra Bfna, *Understanding Resource Types and Traits With RAML,* https://dzone.com/articles/understanding-resourcetypes-and-traits-with-raml (retrieved 26.06.2019)

Vasile ANGHELUȚĂ
"Lucian Blaga" University of Sibiu, Faculty of Sciences
Street Dr. Ion Ratiu no 5-7, 550012 Sibiu
ROMANIA
E-mail: angheluta.vasile@gmail.com

Eduard-Traian ȘTEFĂENSCU
"Lucian Blaga" University of Sibiu, Faculty of Sciences
Street Dr. Ion Ratiu no 5-7, 550012, Sibiu
ROMANIA
E-mail: eduard.traian.stefanescu@gmail.com

Felix HUSAC
"Lucian Blaga" University of Sibiu, Faculty of Sciences
Informatics Street Dr. Ion Ratiu no 5-7, 550012 Sibiu
ROMANIA
E-mail: husacfelix@gmail.com

Motion Transfer using Deep Learning

Andrei-Timotei Ardelean

Abstract

Motion Transfer refers to the transfer of observed actions, motion, dynamics of a particular subject to another entity. It's aim is to replicate and adjust the behaviour to the target object which could have different characteristics. This makes the task challenging as the source movement must be represented in such a way that it can be adapted and reproduced on the target. Nowadays, this topic plays an important role in animations found in cartoons or video games. In this paper the issue is addressed by looking for an automated way of transferring the movement of a human body to another by learning the dynamics of the target body using deep neural networks. A detailed analysis of different existing methods and results on the specific topic is presented along with the improvements and original contributions in our implementation of the technique.

1   Introduction

Older animations, in cartoons for example, were made by drawing each frame by hand, a laborious work that would take a lot of time. Since then, different methods for animating characters, both 2D and 3D have been developed. Most of them require that a model of the target body is created and then motion is represented on this model, allowing it to easily be animated at will. However, it's definitely not easy to build a complex, realistic 3D model of a person for example. The investigated Deep Learning approach tries to skip this step by training a neural network which learns the dynamics of a person from a video stream of the target performing a range of moves.

The trained network is then used to generate video content with the target acting as required. A good performance metric is the model capability to mimic the actions of a person from another video source. Thus, the specific task being solved is to directly output a video of the target, given the frames from some other subject.

2   Related Work

A variety of approaches have been considered to create synthetic content, driven by constraining motion. Older attempts (research done way before the remarkable progress in high quality image synthesis) are mainly based on image construction by using existing frames. In order to generate an image of the target in a specific state, this method firstly searches for images in one or more real recordings of the target for those which closely resemble the desired state. To retrieve this, in Recognizing Action at a Distance [2] a nearest neighbor search from a database of annotated video sequences is used. In Driving visual speech with audio [3] a matching video is determined by computing a triphone distance to each entry in the database. The selected results are then transformed and combined to produce the final output. In approaching the motion retargeting problem in animations, there are also several methods focusing on 3D motion mapping between the subject and the target skeleton. These methods, however, require annotated data with 3D information.

Toward generating images of human in required poses, there are multiple studies, featuring different capabilities and implementation specifics. Most of these studies are using either GANs (Generative Adversarial Networks) or VAEs (Variational Autoencoders). The method described in [4] allows both transfer of appearance and pose and outperforms PG^2[7] in terms of SSIM (Structural similarity); [9] also allows foreground, background, pose manipulation. The proposed method in [8] does not use an indirect representation of the pose which most of the works use, such as [4][5] and [6]. These papers present impressive results, showing the ability to synthesize sharp images even at high resolution. A technique used to achieve super-resolution is well explained in pix2pixHD [12], state-of-the-art method in image to image translation. Using a coarse-to-fine generator trained progressively details can be synthesised in a stable way.

To date, most remarkable results in temporal coherent video generation based on human pose conditioning are vid2vid[6] and [5] which both rely on the aforementioned method to output high quality independent frames, and have different solutions to maintain sensible flow between successive frames. As a result, these papers will be the ones more thoroughly analyzed and used for comparison.

3   Strategy

The first step in generating coherent high quality video of the target performing demanded actions is generating singular frames in a desired state. Recent development in the field of image synthesis promise impressive results on this task. Currently the most prominent approaches are autoregressive models [11], variational autoencoders (VAE)[10] and generative adversarial networks(GAN) [1] all with their strengths and weaknesses [13]

The model build in this work is based on GANs. Typically a GAN architecture consists of two networks: a generator and a discriminator. During training these networks will play a zero-sum game. The generator is tasked to produce synthetic data from latent code so that the discriminator cannot differentiate between real and fake images. Ideally the distribution of generated samples should be indistinguishable from the real images distribution. In order to asses whether the samples come from the fake generator or the training truth samples a critic is trained. The signal from the critic is used to build the loss for the generator which learns to create better images in order to fool the discriminator. The discriminator, in turn, becomes better at distinguishing fakes. As a result both networks are improved in an adversarial way leading to realistic looking images.

Since in our case the generated image must represent a specific state we must use a Conditional GAN, in which the generated output is conditioned on a structured input [14] that comes in addition to the noise distributions. In order to control the output of the generator, a convenient representation of the constraint (the human pose of the subject in our work) must be used.

4   Pose Estimator

As the model must generalize to any subject after training, we must extract an abstract, individual independent representation of the pose from the input images of the subject. Article [5] uses state of the art framework OpenPose [15] for human 2D pose estimation. This method placed first in the inaugural COCO 2016 keypoints challenge, and significantly exceeded previous state-of-the-art results in Multi-Person benchmark; sample output in Figure 1. [6] uses both OpenPose and DensePose for conditioning. "Dense pose estimation aims at mapping all human pixels of an RGB image to the 3D surface of the human body"[16]; sample output in Figure 2.

In our implementation the pose estimation offered by OpenPose was preferred because it facilitates more detailed keypoint detection for hands and face; this enables motion transfer even for face/hand smooth movements.

The output of the OpenPose keypoint detection library comes in two formats: keypoint locations which contain the estimated spatial coordinates of the predetermined 25 key parts of a human body; and an image of the skeleton in which the keypoints are logically linked together. The same for the options to output face and hand details.
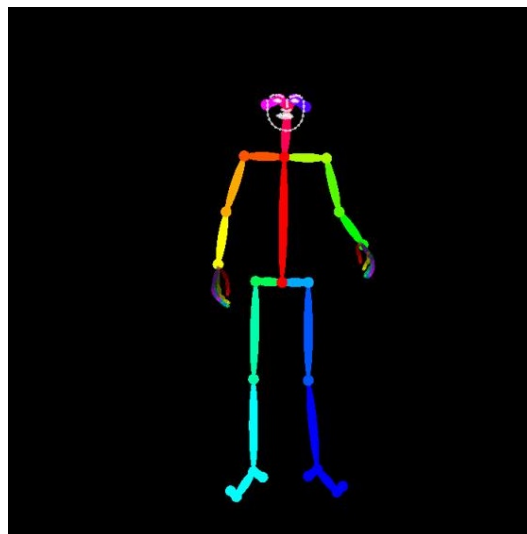
Figure 1: DensePose Example

Figure 2: OpenPose Example

5 Training process

The training data consists of one source video of the target for which we want to learn and model the dynamics. The actions performed in this training video are not restrictive, however it is recommended that a large range of poses and transitions are reached so that when an inference it's required from a novel pose there should exist at least marginally similar instances in the training data. Also it's discouraged to use OpenPose if in most of the training images the target is facing the opposite direction of the camera, as the keypoints are not well detected in this case. The training video is then separated in individual frames which are preprocessed by running the pose estimator on each of them. This creates a new dataset of images that are going to be fed to the GAN architecture along with the real images.

5.1 Individual Frames Setup

For each real frame Y and pose frame X in the preprocessed dataset, a generator G attempts to synthesize an image of the target from pose X alone. Let $\widetilde{X}$ be the output of the generator, so $\widetilde{X} = G(X)$. The discriminator D is then tasked to output whether the image $\widetilde{X}$ is real or fake. A bad image (not looking real) steers the generator to synthesize better images, whilst when the discriminator is fooled it is updated to better discern between fakes produced by G and real images from the training data. Furthermore, the discriminator must check if the image correctly matches the pose, thus it receives as input an (image, pose) pair. In addition to the typical loss for the generator, given by the similarity decided by D, feature-matching loss and VGG loss (Perceptual loss using VGGNet [17]) are also used as proposed in [12]. These losses also appear in the formulation of the objective function in Everybody Dance Now [5] and appear to be of great help both in stability and quality of the end-result. Feature-matching loss is computed by comparing the activations in intermediate layers of the discriminator and VGG loss is the perceptual reconstruction loss [18] which compares features in the layers of a pretrained VGG network.

The discriminator receives 2 pairs((pose, fake gen) and (pose, real)) at each generation step to balance between type I error - classifying a real image as fake and type II error - classifying a fake image as real.

If trained in a progressive manner as proposed in [12] and [13] one can then synthesize from a novel pose real looking, high quality images in this way. However, if successive pose frames from a video are used to generate a video stream the results are not satisfying as the generated frames lack temporal coherence. This causes jittering and breaks the continuity of motion. A naive frame transition smoothing approach based on interpolation or other means would not completely fix the problem as the issue relies not only in the variation of the output given a particular constraint but in the fact that the generator

only captures the state and not the context. Without context, elements like hair movement that depend on the flow, cannot be well represented.

5.2 Temporal Coherence

In order to address this problem, information prior to the current frame must be captured. [5] simply solves this problem by inputting the previous generated frame into the generator. In this scenario the discriminator will also be responsible for determining if an image is coherent with it's history (1 frame context). In this case the input for the discriminator will be a quadruple of (previous pose, current pose, previous generated / real frame, current generated / real frame). The discriminator will receive each time both a fake and a real tuple. Their strategy used in training is to predict two consecutive frames at a time. $G(x_{t-1})$ is conditioned on pose frame x_{t-1} and a zero image. The second output $G(x_t)$ is conditioned on pose frame x_t (pose calculated by the estimator at time t) and the output of the generator on the previous pose $G(x_{t-1})$.

The loss used here is adapted from the original pix2pixHD paper [12]:

$$\min_G((\max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{smooth}(G, D_k))+$$

$$\lambda_{FM} \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k)+$$

$$\lambda_{VGG}(\mathcal{L}_{VGG}(G(x_{t-1}), y_{t-1}) + \mathcal{L}_{VGG}(Gx_t, y_t)))$$

$$(1)$$

Where

$$\mathcal{L}_{smooth}(G, D) = E_{(x,y)}[logD(x_{t-1}, x_t, y_{t-1}, y_t)] + E_x[log(1 - D(x_{t-1}, x_t, G(x_{t-1}), G(x_t)] \qquad (2)$$

and

$$\mathcal{L}_{FM}(G, D) = E_{(x,y)}[\sum_{i=1}^{T} \frac{1}{N_i}(D^{(i)}(x_{t-1}, x_t, y_{t-1}, y_t) - D^{(i)}(x_{t-1}, x_t, G(x_{t-1}), G(x_t))] \qquad (3)$$

The FM loss, is a perceptual loss similar to $\mathcal{L}_{VGG}$ where the features are calculated from the layers of the discriminator. T is the number of layers and N_i is the number of elements in layer i. $D^{(i)}$ is the output of layer i of the discriminator.

Training with this method did not prove very stable in conducted experiments as the quality of the image would rapidly decay after about 200 frames when testing. The reason for this could be that small inaccuracies in synthesizing a frame will be transferred to next one and so forth and the model does not learn to be robust to misformed input and the image quality degrades.

The approach in [6] is more complex. In order to represent the motion transition between two frames it uses an optical flow prediction network. To generate a new image it warps the previously generated image based on the optical flow and then blends in new synthesized pixels by using a 'soft' mask prediction network. These parts are calculated by using the last L frames as a context. After generating a new frame $\widetilde{x}_t$ it used in the prediction for $\widetilde{x}_{t+1}$ and so on, the feed-forward generative network F is trained in a recursive manner.

While the capabilities of the network [6] extends over the needs for motion transfer, the training method observed here can be used in the first scheme analyzed. As a result, in our implementation the approach used is a combination of the two. Since in the final implementation $L = 1$ proved to be enough to get temporal coherence the objective is very similar to (Equation 1), however only a 2-scale discriminator is used and at time t only the perceptual loss for frame t is used.

At time $t = 1$, instead of using a zero image as in [5] a ground truth image $\widetilde{T}$ of the target in a neutral pose is fed into the network as a placeholder for the previous synthesized image. From $t > 1$ input is sequentially fed into the generator as shown in Figure 3 for all the images in the training set. To increase stability, a progressive increase in generated video length was implemented. See section 8 for more details.
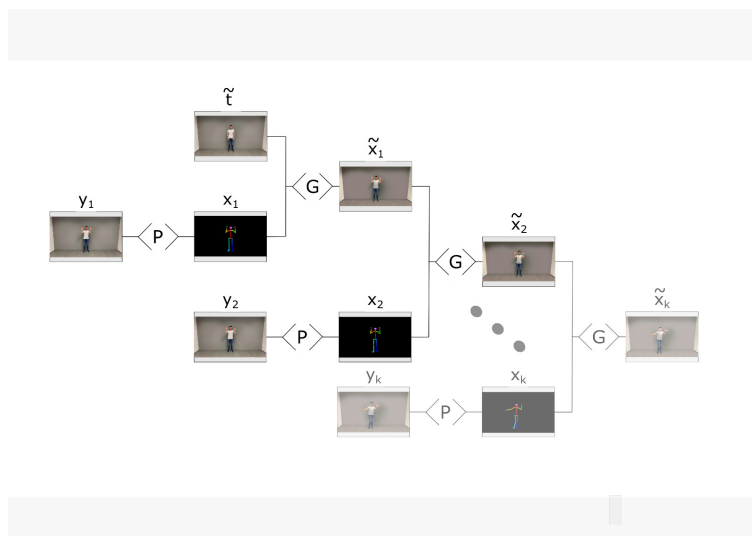
Figure 3: Sequential input into the Generator

6 Novel pose synthesis

Similarly to the training phase, the frames from the source video from which the motion must be transferred are firstly run through the pose estimator. Since in different videos the subject can have different body proportions and / or be at an arbitrary distance away from the camera, the results from the estimator must be aligned with the setup used when recording the target model. While in conducted experiments the network is still able to synthesize modest results without this step the target appears unnaturally taller/shorter or slimmer/plumper than the original. In Everybody Dance Now a normalization step is also applied: Despite using OpenPose the pose frame construction from the joint coordinates in not using the existing functionality instead drawing lines between connected joints themselves. Prior to the drawing, the joints are transformed. "We find this transformation by analyzing the heights and ankle positions for poses of each subject and use a linear mapping between the closest and farthest ankle positions in both videos. After gathering these statistics we calculate the scale and translation for each frame based on its corresponding pose detection."[5]. The method used, however, implies calculating the median ankle position so an offline look up through the subject video is required. In our implementation the scale and translation are calculated based on a single, initial frame of the source video aligned to the $\widetilde{T}$ image (in a neutral pose) from the target video. By using this method the video generation is online, making it possible to achieve live video translation given enough computational power.

7 Face Enhancer

Both [5] and [6] allow special enhancements for face generation, [5] by creating a specialized Face GAN and [6] by providing an option to add another discriminator that works on the face region. The Face GAN is built upon the global generator of pix2pixHD [12] and works by generating a region around the face, being given as input the same region from the pose and the synthesized image by the main generator G, respectively. It was shown to produce improved results adding convincing details and facial features.

The residual output of the Face GAN is added to the original face region to obtain the final result. In our experiments when this method is used, the Face GAN would usually not converge to good results as it would not learn to adapt to the noise in the image produced by the main generator. If a regular synthesis is used (not residual) it successfully learned to correct the face details in a stable way.
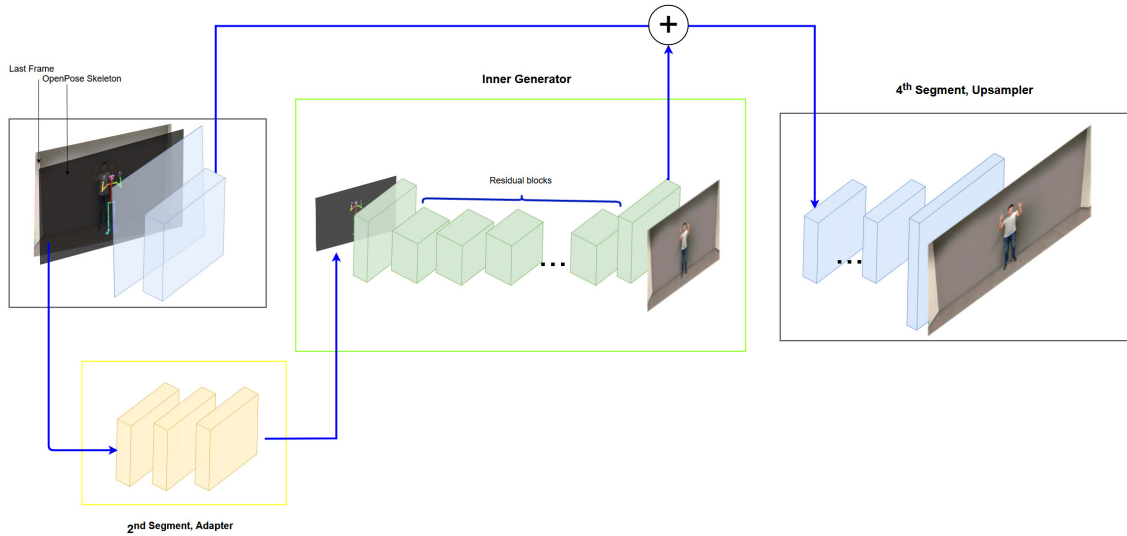
Figure 4: Neural Model Architecture

8 Experiments

One aspect worth mentioning regarding single frame generation is a problem noticed when using pix2pixHD architecture. When the pose skeleton occupied a small part of the image, the conditioning information would get lost and the parts of the image closer to the edge were very blurry. This issue was solved by adding some spatial information to the pose frame, that is a thick border close to the margins of the image.

Something that proved very effective in increasing stability and reducing training time was a progressive increase of synthesized video length. As mentioned, in the temporal coherent setting, the generator uses as input at time t, along pose frame x_t, the image generated at time $t - 1$, that is $\widetilde{x}_{t-1}$. So $\widetilde{x}_t = G(x_t, G(x_{t-1}, G(x_{t-2}, ...G(x_0, \widetilde{T}))))$. Since every frame depends on everything generated before it, inaccuracies get propagated and eventually the quality degrades entirely. When trained progressively, we keep the dependency chain smaller until the networks learns to make fewer mistakes. To do this, with a probability of ϵ, the history is reset, meaning that we input a real frame to the generator: $\widetilde{x}_{t-1} = G(x_t, y_{t-1})$ or $\widetilde{x}_{t-1} = G(x_t, y_{t-1}, y_{t-2}, ..., y_{t-L})$, if $L > 1$. To the same end, to get a model more robust to input defects, small noise is added to the input. In practice the L+1 images used as input for the generator are concatenated on the channel axis and each pixel in each channel is altered individually by a small factor.

9 Neural Model Architecture

As captured in Figure 4 the pipeline (Face Enhancer is not included) consists of 4 segments. To simplify the figure L is considered equal to 1. In the first part, the image is preprocessed to obtain the pose estimation. The result is concatenated with the previous L generated frames then run through an adapting segment that prepares the input for the inner Generator. The features from the layer before the last of the inner generator which was pretrained to output images at a smaller resolution are used. These features are then processed in the forth and final segment which outputs the synthesized image $\widetilde{x}$ of the target.

When the Face Enhancer is used, after the main network outputs $\widetilde{x}$ a small region around the face is run through this second Generator. The corrected region is then interpolated back into $\widetilde{x}$ resulting in the final output.
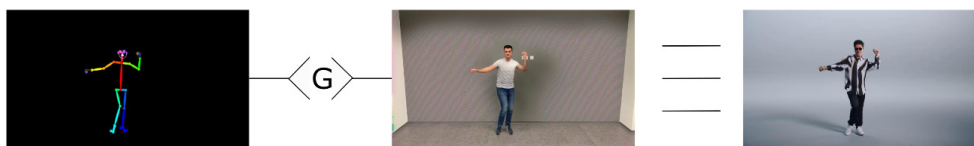
Figure 5: Novel Pose Generation Example

10   Implementation details

In the final implementation ϵ, the real image sampling probability is kept with a value of 0.5 for the first two epochs, and then annealed to zero over eight epochs. The inner generator, (global generator as named in pix2pixHD) is trained for 10 more epochs. After the first resolution is converged the full (main) network is trained. The inner network is freezed and ϵ is reset to 0.5 following the same policy. At epoch 10, ϵ reaches zero and the global generator is unfreezed, and after 10 more epochs the generator usually stops improving. The Face Enhancer network receives a 128x128 rectangular crop from the image synthesized by the main network, centered at the face. This network usually learns pretty fast and can be either trained along the main network, or afterwards for about 2-3 epochs.

Good results are obtained using a regular 10 minutes HD video at 30fps with the target performing more or less random moves. [5] uses 20 minutes of video at 120fps and also pre-smooth pose keypoints in source videos which was not not necessary in our case.

The training took about 3 days for the inner generator at 512p resolution. Another 2 days were necessary for the full network to converge. The Face Enhancer needed just a few more hours after everything else was trained. This time is recorded when training on an NVidia RTX 2080 Ti GPU. Adam optimizer was used with a learning rate of 0.00008 and $(\beta_1, \beta_2) = (0.5, 0.999)$

11   Results

Overall, the results are of sensible quality as the model can generate long videos, preserving temporal coherence and without quality degradation over time. The movement imitation is fairly precise and individual frames are sharp and realistic (though dependent, of course, of the precision of the estimated pose). The final results of the implemented model can be seen in Figure 5. The Face enhancer successfully increases the quality of the selected region as seen in Figure 6. There is work in progress to compare the results against vid2vid in terms of SSIM and FID (Fréchet Inception Distance) [19]. A fair comparison against [5] it's not possible as the authors did not release their source code to the public.

12   Conclusions

As shown, multiple researches are improving upon pose detection, pose conditioned image synthesis and also video generation using various techniques. The results are not perfect as the generated video are not yet indistinguishable to real ones by a human observer. However, they show promising quality at impressive resolution. The motion is correctly transferred to the target without the need of specific annotated data. A single not restrictive video is required. There is progress to be made toward the speed of the image synthesis. The pose estimator OpenPose claims real time capability so real time motion transfer is not far fetched. Training time can also be considered an issue, a problem which this paper approaches by using the progressive video length generation.
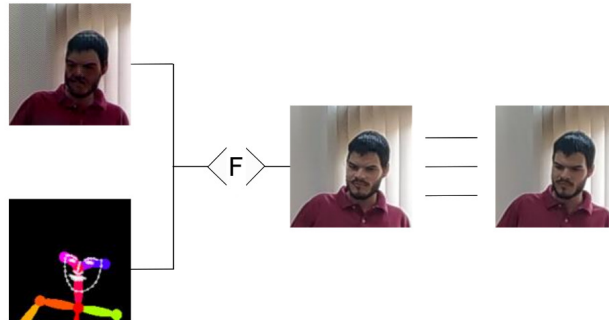
Figure 6: Face Enhancer

References

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, *Generative Adversarial Networks* arXiv:1406.2661

[2] Alexei A. Efros, Alexander Berg, Greg Mori, Jitendra Malik, *Recognizing Action at a Distance* IEEE International Conference on Computer Vision 726733

[3] Christoph Bregler, Michele Covell, Malcolm Slaney, *Video rewrite:Driving visual speech with audio* Proceedings of the 24th annual conference on Computer graphics and interactive techniques 353360, 1997

[4] Patrick Esser, Ekaterina Sutter, Bjrn Ommer, *A Variational U-Net for Conditional Appearance and Shape Generation* arXiv: 1804.04694

[5] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros, *Everybody Dance Now* arXiv:1808.07371

[6] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, Bryan Catanzaro, *Video-to-Video Synthesis* arXiv:1808.06601

[7] Liqian Ma, Xu Jia, Qianru Sun, Bernt Schiele, Tinne Tuytelaars, Luc Van Gool, *Pose Guided Person Image Generation* arXiv:1705.09368

[8] Donggyu Joo, Doyeon Kim, Junmo Kim, *Generating a Fusion Image: One's Identity and Another's Shape* arXiv:1804.07455

[9] Ma, Liqian and Sun, Qianru and Georgoulis, Stamatios and Van Gool, Luc and Schiele, Bernt and Fritz, Mario, *Disentangled Person Image Generation* IEEE Conference on Computer Vision and Pattern Recognition, 2018

[10] Diederik P Kingma, Max Welling, *Auto-Encoding Variational Bayes* arXiv:1312.6114

[11] Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, *Pixel Recurrent Neural Networks* arXiv:1601.06759

[12] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, Bryan Catanzaro, *High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs* arXiv:1711.11585

[13] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, *Progressive Growing of GANs for Improved Quality, Stability, and Variation* arXiv:1710.10196

[14] Mehdi Mirza, Simon Osindero, *Conditional Generative Adversarial Nets* arXiv:1411.1784

[15] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields* arXiv:1611.08050

[16] Rza Alp Gler, Natalia Neverova, Iasonas Kokkinos, *DensePose: Dense Human Pose Estimation in the Wild* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7297-7306, 2018

[17] Karen Simonyan, Andrew Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition* arXiv:1409.1556

[18] Justin Johnson, Alexandre Alahi, Li Fei-Fei, *Perceptual Losses for Real-Time Style Transfer and Super-Resolution* arXiv:1603.08155

[19] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter, *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium* arXiv:1706.08500

Andrei-Timotei ARDELEAN
Transilvania University of Brasov
Faculty of Mathematics and Informatics
Iuliu Maniu, no 50
ROMANIA
E-mail: *timitoc@yahoo.com*

An Open-Source Approach to Prepare Scanned Images for OCR

Tiberiu Bujor

Abstract

The progress made by the big companies in the artificial intelligence domain is often shared with the public as open source or by allowing free use of software programs over the internet. Google drive OCR removes the burden of having to obtain data needed to feed a classifier by using its huge dataset. The greatest results are always obtained when the correct information is provided and, using Google's OCR is no different because image repair is a resource costly process and should be done separately if one is to obtain the best results. This paper describes methods and algorithms developed for document image enhancement and retrieval that could be used before the optical character recognition process.

1. Introduction

Document images usually suffer from several defects mainly due to natural ageing, environmental conditions, usage, poor storage conditions of the original, as well as human manipulations during the scanning process. Since document image analysis and recognition methods usually assume a smooth background and a good quality of printing or writing, it is imperative to have an efficient image enhancement process in order to restore the good quality of document images. Moreover, document image enhancement process enhances the readability of text areas and permits less image storage space.

Google Drive's Optical Character Recognition (OCR) lets you convert PDF files with images with text into text documents with support for over 200 different languages. The major disadvantage is that it's limited when it comes to resolving problems mentioned previously, that are obtained from low contrast and uneven background illumination, bleed-through, shining, or shadow-through effects, damaged characters or noisy background, borders or parts of neighboring page, etc.

This paper is organized as follows: section 2 provides the general techniques for document image enhancement and normalization, section 3 will review the most popular and current algorithms used in document image enhancement and normalization applications, and section 4 concludes the paper.

2. Enhancement techniques

Image enhancement can be easily described as transforming an image f into image g using T. (where T represents the transformation). The main categories in which these techniques fall are either of the Spatial domain (pixel oriented) or based on a Frequency domain (Fourier transform).

One major issue in image processing is *contrast enhancement*, to solve the low contrast and uneven background illumination of an image, one must focus on enhancing the luminosity between background and text areas, smoothing of background texture, as well as the elimination of noisy areas.

Some popular methods in this manner are:

- Histogram equalization (most simple, and effective)
- Otsu's thresholding [1]
- Kapur's entropy [2]
- Kittler and Illingworth minimal error [3]
- Leung generalized fuzzy operator preprocessed with histogram equalization and partially overlapped sub-block histogram equalization (achieves good performance when working with extremely low contrast and low illuminated document images) [4]
- Nomura method in which morphological operations are used to remove undesirable shapes called critical shadows on the background of document images before proceeding to binarization [5]

One other major issue consists of the *Bleed-Through, Shining, and Shadow-Through Effects* that can be categorized into either non-registration (or blind) or registration (or non-blind) methods, depending on whether the verso image of a page is available and precisely registered to the recto image.

a) When the two sides of a page are treated independently and processed separately (*non-registration-based methods*). Those methods attempt to clean the front side of a document without referring to the reverse side. Most of these methods treat bleed-through interference as a kind of background artifacts or noise and remove it using threshold-like techniques.

b) When the verso image of a page is available and precisely registered to the recto image (*registration-based methods*). In order to register the verso with the recto image, several techniques have been proposed based on:

    (i)   intensity characteristics, intensity patterns are compared using correlation metrics;

(ii)   feature characteristics, correspondence between features such as points, lines, and contours is searched;

(iii)   transformation models;

(iv)   spatial and frequency domain.

When character areas are not visible or there is noise in the background (*noise reduction method*) common operations are applied in order to enhance the quality of a binary document image; these include masks, connected component analysis, morphological operations, as well as shrink and swell operations.

Enhancement of grayscale or color images has proved to be efficient by application of Ajayi's Partial Diffusion Equation (PDE) for enhancing text in degraded document images. His algorithm is based on a lookup table classification algorithm that learns the corrections of patterns of text degradation in document images.[11]

When borders or parts of adjacent pages are visible in the current image, methods that can be applied are Connected Component Analysis-Based Methods, Projection Profile-Based Methods and "Flood-Fill" Algorithms.

When document digitization is done with flatbed scanners or camera-based systems, the resulting images can often suffer from skew, warping, and perspective distortions. The scanned text image may happen to be rotated by half or even be upside down. A document image normalization step is imperative in order to restore text areas horizontally aligned without any distortions as well as in a straight angle.

Portrait/landscape orientation detection is mainly accomplished by using projection histograms as well as by counting black-to-white transitions.

When the skew angle ranges between -10° to 10° we can apply deskew and deslant methods which fall into the following categories:

a) *Projection Profile-Based Skew:* works by calculating a series of horizontal projection profiles as we rotate the document page at a range of angles. The optimization of an objective function for a given skew angle leads to the actual document skew.

b) *Hough Transform-Based Skew:* Each black pixel is mapped to the Hough space (*p, 0*) and the skew is estimated as the angle in the parameter space that gives the maximum sum of squares of the gradient along the *p* component.

c) *Nearest-Neighbor Clustering-Based*: spatial relationships and mutual distances of connected components are used to estimate the page skew. The direction vector of all nearest-neighbor pairs of connected components is accumulated in a histogram and the peak in the histogram gives the dominant skew.

d) *Cross-Correlation:* are based on measuring vertical deviations among foreground pixels along the document image in order to detect the page skew.

Using a flatbed scanner or a digital camera also results in several unavoidable image distortions due to the form of printed material, the camera setup, or environmental conditions (page erosion).

Many different techniques for dewarping can be classified into two main categories based on 3-D document shape reconstruction and 2-D document image processing.

a) Extraction of the 3-D information of the document and then models the page surface by curved developable surfaces to estimate the 3-D shape of the page using texture flow fields.

b) Detection of distorted text lines at the original document image which is a well known hard task. Some of these techniques propose a method to straighten distorted text lines by fitting a model to each text line.

3.   Enhancement algorithms

3.1 Detect and fix skew in images containing text [6]

Methods used: *Canny Edge Detection, Hough Transform*

How it works:

- Converts the image to grayscale
- Performs Canny Edge Detection on the Image
- Calculates the Hough Transform values
- Determines the peaks
- Determines the deviation of each peaks from 45 degree angle
- Segregates the detected peaks into bins
- Chooses the probable skew angle using the value in the bins

Technical specifications:

- Year: 2017
- Programming
  Language: Python
- License: MIT
- Author: Kakul Chandra

Installation:

1. # mkdir deskew-test
2. # cd deskew test
3. # mkvirtualenv .
4. # pip install alyn
5. # source bin/activate

Requires:

- numpy
- matplotlib
- scipy
- scikit-image

Usage:

```
from alyn import Deskew
d = Deskew(
input_file='path_to_file', display_image='preview the image on screen',
output_file='path_for_deskewed_image',
r_angle='offest_angle_in_degrees_to_control_orientation')
d.run()
```

```
./skew_detect.py -i image.jpg
```
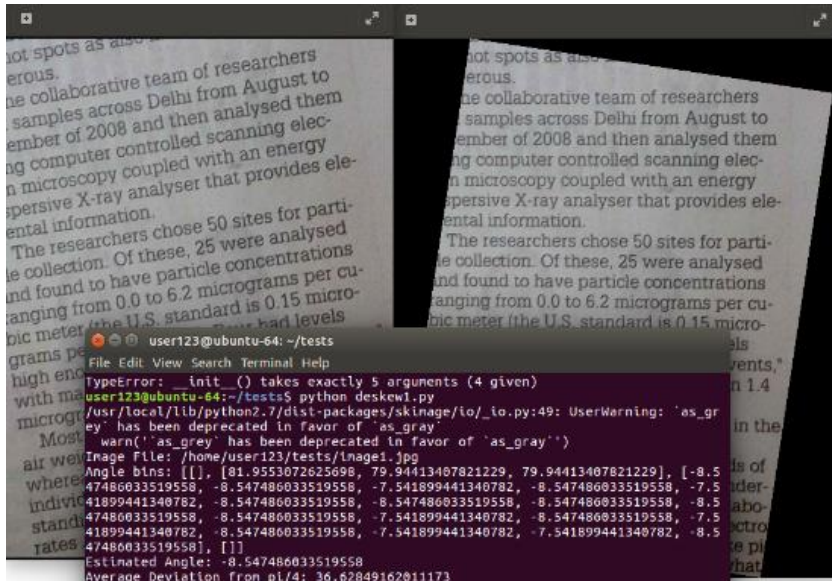

Before (left) and after (right) of test run with angle < 5 degrees [Fig. 1]

3.2 Contrast Enhancement for low-light images [7]

Methods used: *Exposure Fusion Network*

How it works:

- design the weight matrix for image fusion using illumination estimation techniques
- use camera response model to synthesize multi-exposure images
- find the best exposure ratio so that the synthetic image is well-exposed in the regions where the original image was under-exposed
- input image and the synthetic image are fused according to the weight matrix to obtain the enhancement result

Technical specifications:

- Year: Sep 24, 2018
- Programming Language: Python
- License: Open
- Author: Andy Huang

29

Installation:

    # mkdir enhancement-test

    # cd enhancement-test

    #   git   clone https://github.com/AndyHuang1995/Image-Contrast-Enhancement

Requires:

- numpy
- scipy
- scikit-image
- cv2
- imageio
- matplotlib
- git
- skimage

Usage:

```
# python ying.py <input image>
```
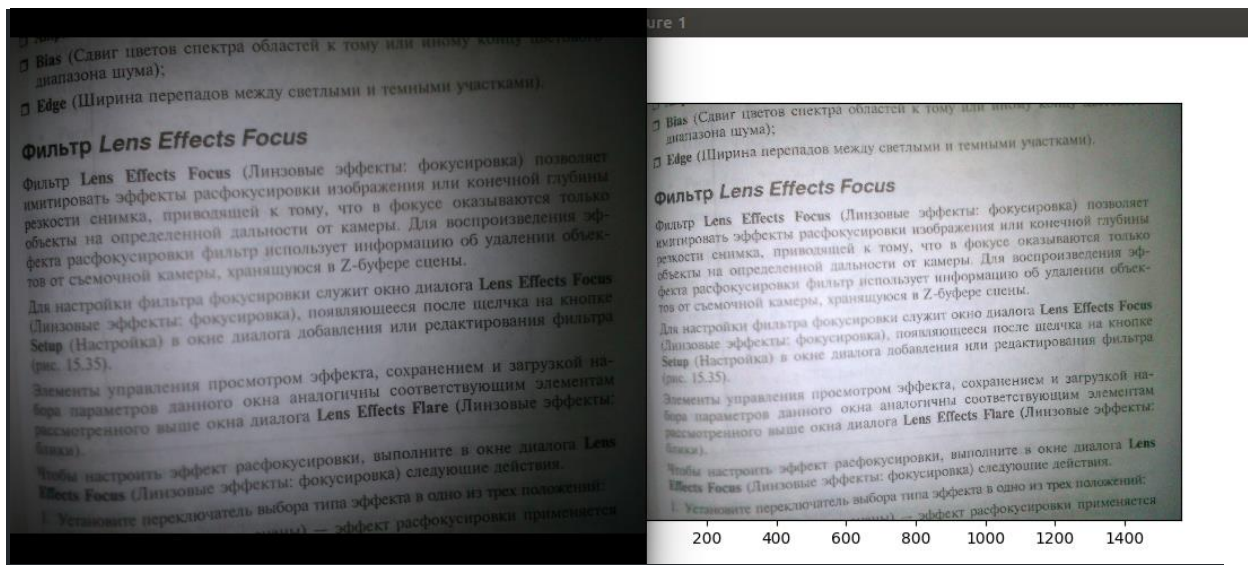

Before (left) and after (right) of contrast enhancement algorithm with no input parameters [Fig. 2]

3.3 Image Shadow Detection and Removal [8]

Methods used: *Processing in HSV color space, histogram matching*

How it works:

- take the derivatives of the original image
- apply a mean shift segmentation
- setting shadow edges derivatives to 0 and reintegrate the image by:
  - solving a Poisson equation, 2-dimensional integration
  - generating random Hamiltonian paths, 3-dimensional integration

Technical specifications:

o   Year: Nov 21, 2017
o   Programming
Language: Matlab
o   License: Open
o   Author: Jiarui Gao

Requires:

- Git
- Matlab

Installation:

install matlab & plugins

mkdir shadow_remove-test

cd shadow_removetest

git clone https://github.com/kittenish/Image-Shadow-Detection-and-Removal

Usage:

 # open main.m in matlab and select the image
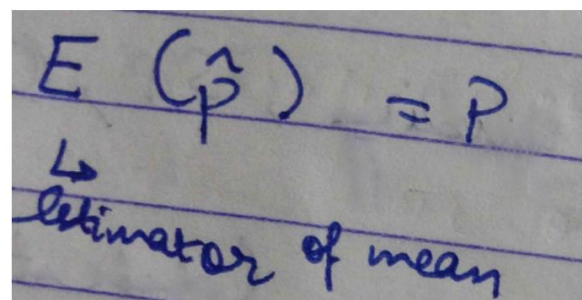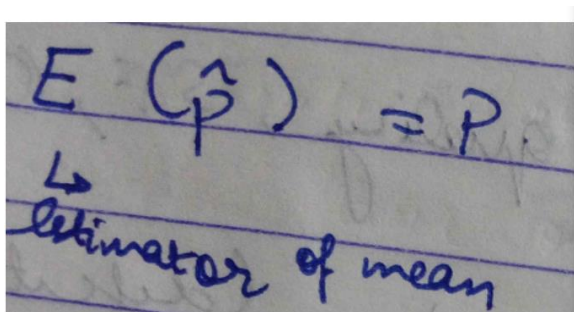
Before (left) and after (right) of the shadow remove algorithm on a selected image. [Fig. 3]

3.4 Denoising [9]

Methods used: *Despeckle and Enhance*

How it works:

- consider each pixel in the image
- sort the neighboring pixels into order based upon their intensities
- replace the original value of the pixel with the median value from the list

Technical specifications:

- Year: December 15, 2018
- Programming Language: Bash
- License: For personal use only
- Author: Fred   Weinhaus

Requires:	Installation:

- Git
- Matlab

```
# mkdir noisecleaner-test
# cd noisecleaner-test
#wget
http://www.fmwconcepts.com/imagemagick/downloadcounter.php?sc
riptname=noisecleaner&
dirname=noisecleaner
# chmod +x *
```

Usage:

```
# ./noisecleaner-m 2 -n 10 -f all image.png image2.png
```
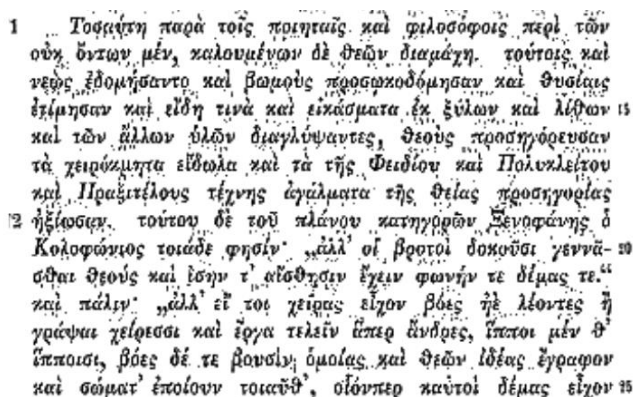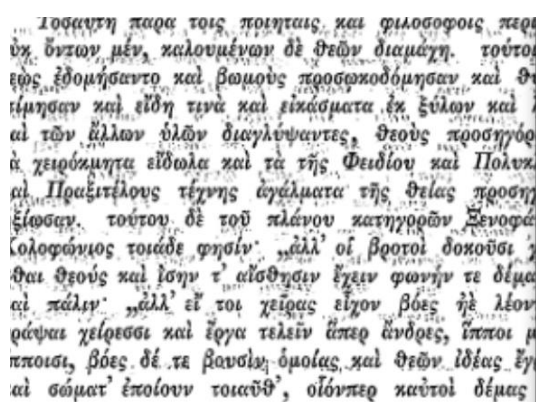

Before (left) and after (right) of the noise removal algorithm on a text image. [Fig. 4]

3.5 Dewarping [10]

Methods used: *Split the text into lines and find a warp or coordinate transformation that makes the lines parallel and horizontal*

How it works:

- Obtain page boundaries
- Detect text contours
- Assemble text into spans
- Sample spans
- Create naive parameter estimate
- Remap image and threshold

Technical specifications:

- Year: Oct 2, 2016
- Programming Language: Python
- License: MIT
- Author: Matt   Zucker

Installation:

Requires:

mkdir dewarp-test

- scipy
- opencv3

cd dewarp-test

- pillow
- git

git clone https://github.com/mzucker/page_dewarp

Usage:

```
# python python page_dewarp.py image1.jpg
```
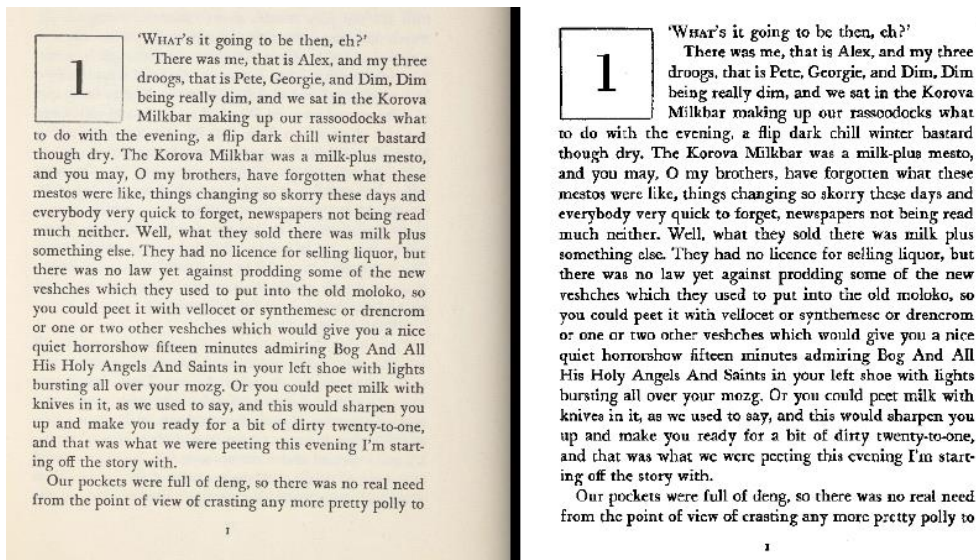

Before (left) and after (right) of the dewarp algorithm on a book page image. [Fig. 5]

4.   Conclusion & Future improvements

The discussed techniques are more than likely to have an open source implementation that is relatively easy to find and can be used with little to no configuration required. The highlighted algorithms performed better than expected and could be integrated in a automated tool (preparator) that would further serve them to the OCR. Another approach would be to look at the source code of a bigger tool that has all these features already integrated (GIMP) and use the code in a specific implementation targeted to only fix scanned text images.

Installing the necessary libraries for running these algorithms is a very straight forward process that should not require a lot of time for an experienced user. I used an Ubuntu 16.04 VM to test each algorithm and wrote a few lines of Python to fix some incompatibility issues.

If one individual confronts with the problem of extracting text from an image where the OCR software is not capable of detecting characters, a simple revision of the presented algorithms can come in very handy and for further approach methods he can study the standard presented techniques and search for a common implementation.

References

[1] Otsu N, *A threshold selection method from Gray-level histograms,* 1979 – https://ieeexplore.ieee.org/document/4310076

[2] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, *A new method for gray-level picture thresholding using the entropy of the histogram,* Computer Vision, Graphics, and Image Processing, Vol. 29 – https://www.sciencedirect.com/science/article/pii/0734189X85901252

[3] J. Kittler, J. Illingworth*, Minimum error thresholding,* Pattern Recognition, Vol. 19, No. 1, 1986 – https://www.sciencedirect.com/science/article/abs/pii/0031320386900300

[4] Leung CC, Chan KS, Chan HM, Tsui WK*, A new approach for image enhancement applied to low-contrast–low-illumination IC and document images*, 2005. – http://hub.hku.hk/handle/10722/73568

[5] Nomura S, Yamanaka K, Shiose T, Kawakami H, Katai O*, Morphological preprocessing method to thresholding degraded word images,* 2009. – https://www.sciencedirect.com/science/article/abs/pii/S016786550900049X

[6] *Deskew algorithm* – https://github.com/kakul/Alyn

[7] *Contrast Enhancement algorithm* – https://github.com/baidut/OpenCE

[8] *Shadow Removal algorithm* – https://github.com/kittenish/Image-Shadow-Detection-and-Removal

[9] *Noise cleaner algorithm* – http://www.fmwconcepts.com/imagemagick/noisecleaner/index.php

[10] *Dewarp algorithm* – https://github.com/mzucker/page_dewarp

[11] *Obafemi-Ajayi T, Agam G, Frieder O (2010) Historical document enhancement using LUT classification.* – https://link.springer.com/article/10.1007/s10032-009-0099-3

Tiberiu BUJOR
"Vasile Alecsandri" University of Bacău
Faculty of Sciences
Bacău ,Calea Mărăşeşti 157
ROMANIA
E-mail: bujor.tiberiu@gmail.com

Job scheduling optimizations in CloudSim
by Monte Carlo method

Dejan Bulaja, Kristina Bozic, Nikola Penevski

Abstract

Cloud computing is a subscription-based service where networked storage space as well as other computer resources can be acquired. Cloud computing is highly desirable in the rapidly growing world of computer technology and because of that the Cloud Computing and Distributed Systems (CLOUDS) Laboratory built the CloudSim framework. It is a very popular open source cloud simulator among the researchers and students. CloudSim represents a customizable framework that allows for uninterrupted modeling and simulation of cloud computing infrastructure. Its usefulness is, however, that it opened up the possibilities of evaluating the hypothesis in a controlled environment. In this environment, experimental results can be achieved easily with all the available customisations that framework has to offer. In this paper, CloudSim's default job scheduling algorithm is modified with the Monte Carlo method with the purpose of increasing the efficiency and utilization of the cloud making the user submitted jobs execute faster producing cost and time savings.

1   Introduction

Cloud computing is the on demand availability of computer system resources, most notably data storage and computing power. The popularization of the term "cloud computing" can be traced back to 2006 [1], when large companies such as Amazon and Google began using it and making patents [2] to describe the new way of accessing software, computing power and files over the web. The coinage of the term can be traced back to 1996 , when in the Compaq Computer Corporation's business plan was stated : "The emergence of the internet is driving the migration of communication and collaboration applications into the internet „cloud"." [3].

CloudSim is a framework which supports modeling and simulating of the cloud environment. For developers it is a tough job to evaluate the performance of their application for different load configurations and settings. By using CloudSim, developers can test their application in controlled and repeatable environment, before deploying it in actual cloud.

Main features of CloudSim are:

1. Support of modeling and simulation of computing environment.
2. A self contained platform for modeling clouds, service brokers, provisioning and allocation policies.
3. Support for simulation of network connections among the simulated system elements.
4. Facility for simulation of federated cloud environment, that inter-networks resources from both private and public domains.
5. Help in the creation and management of multiple independent and co-hosted virtual services on a host.
6. Flexibility to switch between space shared and time shared allocation policies.

2   Background

Cloud computing is affordable, scalable and available on demand to its customers. It is a subscription-based service where data storage and computing power can be acquired [4]. Cloud computing today consists of three types of computing services delivered remotely to clients through the internet. In the near future we might see a change as Buuya mentions [5]. Cloud computing can be defined as "a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resourced based on service-level agreements established through negotiation between the service provider and consumers" [6].

Many companies have become global and are delivering services from the cloud. They use a variety of techniques to meet their consumption demands. Efficency is one of their top priorities [7]. Some notable examples include companies such as Google, Amazon, Microsoft with their well regarded infrastructures including Amazon Web Services, Microsoft Azure and Google Cloud Platform.

The National Institute of Standards and Technology (NIST) defines cloud computing as: a pay-per-use model for enabling available, convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [8].

2.1 Cloud Computing Models

Depending on the customer''s requirements, how they plan to use the cloud computing services there are different business models [9]. They can be grouped in three categories, which are listed below, although there are good arguments showing that this might not be the ideal way to classify them [10]:

1. Software as a Service (SaaS) - Customers purchase the ability to access and use an application or service that is hosted in the cloud. This model offer a complete application to its customers. Also, this model is specific because a single instance of the service is running on the cloud and multiple end users can be served simultaneously
2. Platform as a Service (PaaS) - Customers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating

systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed, but the utility is that the customer can build his own application which runs on the provider‟s infrastructure.

3. Infrastructure as a Service (IaaS) - Customers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure.

2.2 Deployment models

Enterprises can deploy applications on one of the three clouds, and those are Private, Public and Hybrid clouds.

Private cloud is cloud infrastructure operated solely for a single organization or enterprise [11]. The main goal for private cloud is to give their customers more secure environment and to offer greater control over the infrastructure.

Public cloud is considered public when the services are rendered over a network that is open for public use. They are owned and operated by a third party, and all customers share the same infrastructure pool which limits the security and available configuration for each customer.

Hybrid cloud is a composition of both cloud infrastructures addressed above, and it is offering the benefits of both deployment models. By allowing workloads to move between private and public clouds as computing needs and costs change, hybrid cloud can give businesses greater flexibility.

3   CloudSim Architecture
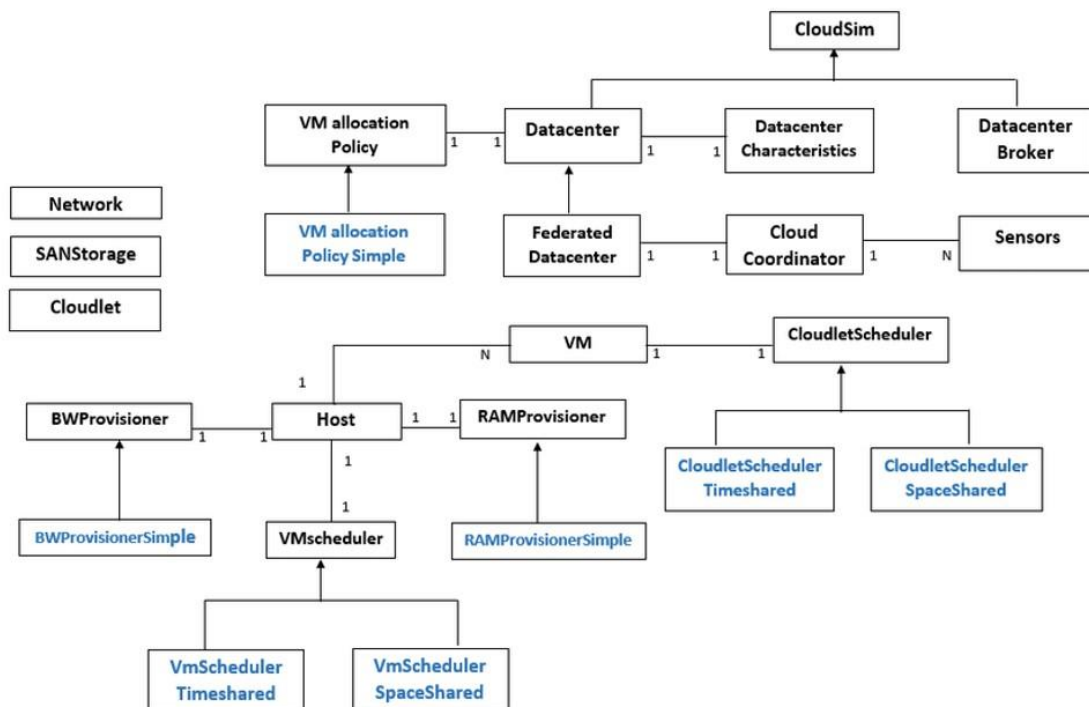
Fig. 1: CloudSim Diagram

CloudSim is a library for the simulation of all kinds of cloud scenarios. It provides essential classes for describing data centres, hosts, virtual machines, cloudlets, and various policies for the management such as scheduling and provisioning. An overview of the CloudSim architecture is shown in Figure 1. We will go over some of the relevant parts for this work.

3.1 Datacenter

Datacenter models the infrastructure services provided by cloud service providers. It consists of multiple computing hosts that model physical resources.
It is described by the following parameters:
1. Total available processing power in MIPS (Million Instructions Per Second)
2. Total available RAM in MB (Random Access Memory)
3. Total storage represented in MB
4. Total bandwidth in mbps.

3.2 Virtual machines

Virtual machine is an emulation of a computer system. Every VM that is created by CloudSim is allocated to a Host where all its computational requirements are met. VM Scheduler is responsible for allocating processor cores to VMs.
VM is described by the following properties:
1. CPU power represented in MIPS
2. Number of processing cores
3. Amount of RAM memory
4. Size of the virtual disk image
5. Bandwidth in mbps
6. Name of the virtual machine

3.3 Cloudlets

Cloudlet (cloud task) is a set of user requests. Cloudlet Scheduler"s function is to assign cloudlets to virtual machines.
 Cloudlet has the following attributes:
1. ID of cloudlet that is used by the broker to assign them to a virtual machine.
2. Length of a task that is to be executed represented in bytes.
3. Output file size is the size of the cloudlet after executing.
4. Number of processors that are required to run the cloudlet.

Both VM Scheduler and Cloudlet Scheduler come with two different allocation policies: Timeshared and Spaceshared, as well as the option to implement user"s own solution.. In the time shared policy Cloudlets are run simultaneously, while in the space shared policy VMs will execute cloudlets in a chronological order, one by one, using full capacity for each cloudlet.

4 Monte Carlo methods and Job Scheduling

4.1 Monte Carlo

Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to make numerical estimations of unknown parameters. They allow for modeling of complex situations with many different variables. The concept was invented by Stanislaw Ulam, a mathematician from the 20th century who devised these methods and used the tools of random sampling and inferential statistics to model likelihoods of outcomes, originally applied to a card game called Monte Carlo Solitaire.

There is no consensus on how Monte Carlo should be defined. Sawilowsky [12] distinguishes between a simulation, a Monte Carlo method, and a Monte Carlo simulation: a simulation is a fictitious representation of reality, a Monte Carlo method is a technique that can be used to solve a mathematical or statistical problem, and a Monte Carlo simulation uses repeated sampling to obtain the statistical properties of some phenomenon or behavior.

These methods are often used today in physical and mathematical problems where it's impossible to use deterministic approaches. This is the case with NP problems where increasing the complexity of a problem linearly requires a solution that has a nondeterministic polynomial time [13]. The number of possible solutions becomes large to search deterministically so we are forced to used methods that use stochastic methods to search the random parts of the total search space in order to obtain results that are close to optimal.

4.2 Using Monte Carlo method for the optimization of Job Scheduling

VMs and Cloudlets are defined by many parameters that affect their speed of execution. The standard scheduling algorithm in CloudSim is round-robin which employs time-sharing. This algorithm is less than ideal since it does not use any type of optimization. Optimizing the order in which Cloudlets execute and to which VM they get assigned would decrease the time that is required to run all Cloudlets. Using Monte Carlo method we achieve this optimization by searching through the space of possible Cloudlet orders in a random manner.

By searching through permutations statistically we can always find an order that would outperform the mean time without optimization. This is done by running the simulation with different permutations of the Cloudlets, recording and observing the results. After observing the times the permutation that has the best result is taken as the solution. What's worth noting is that the results form a Gauss curve. This is due to to the sheer number of parameters that have an effect on the running time. By choosing from a number or permutations, the most likely outcome is the mean value. Chances that the Cloudlet order becomes favourable for us in a way that they utilize the most resources all the time is small which corresponds to to the very small percentage of solutions that are desired.

5 Results and Program Code

5.1 Results

Our example consists of 40 Cloudlets that should be executed on 20 VMs concurrently. For the purpose of testing the implementation all Vms and Cloudlets have different specifications (time that they run, RAM used, storage needed for the output, and bandwidth). We are aiming to optimize the time of execution in such a way that all tasks finish as soon as possible. This is achieved by optimizing the order in which Cloudlets are run.

An optimal run is one where all VMs are fully utilized and all finish at the same time. This guarantees the fastest run time possible. The time when the last VM finished executing the last Cloudlet is called fininshingTime. Each Cloudlet can be executed by any of the 20 VMs. There are 40 Cloudlets so that the number of possible permutations is 40^20 which is 1.09e+32. The search space is too large to be searched fully in a reasonable amount of time. For modern computers to search a space this large would take many magnitudes longer than to execute the Cloudlets in any given order.

This is not practical so we must turn to stochastic methods which search parts of the search space hoping to find a good enough solution. The time that is used to optimize the order in which the Cloudlets execute should be negligible when compared to the time of Cloudlets executing. Depending on which optimization algorithm is run and how long it runs for, it will find better and better solutions for the given problem.

In our example we have used the Monte Carlo method with 500 epochs. We chose this number of epoch empirically as it gives the best ratio of time run and results given. Running more epochs doesn't yield much better results, but takes more time. Running less epochs has a greater chance of finding a good enough solution.

Since no two Cloudlets or VMs are the same there is only one best solution that takes the least amount of time to run. Our implementation of the Monte Carlo method compares the different order of Cloudlets with the time that they would take to execute. After running 500 epochs we declare the the run with the least time the best solution. A visual representation of the results is shown below in Figure 2.

After running our code we take all the results from the Monte Carlo method as a given array of 500 elements: [9.446, 9.403, 8.129, …, 13.348, 8.062, 11.398] and extract the lowest time to find the shortest time of execution of running our simulation which is 6.794ms. The order of Cloudlets that gives this number is the best solution that we managed to find.

The mean time that can be expected from running the same example on a standard, non optimized version of CloudSim is 9.862ms. We have managed to increase the efficiency of the cloud by 45,16 percent by using the Monte Carlo method.
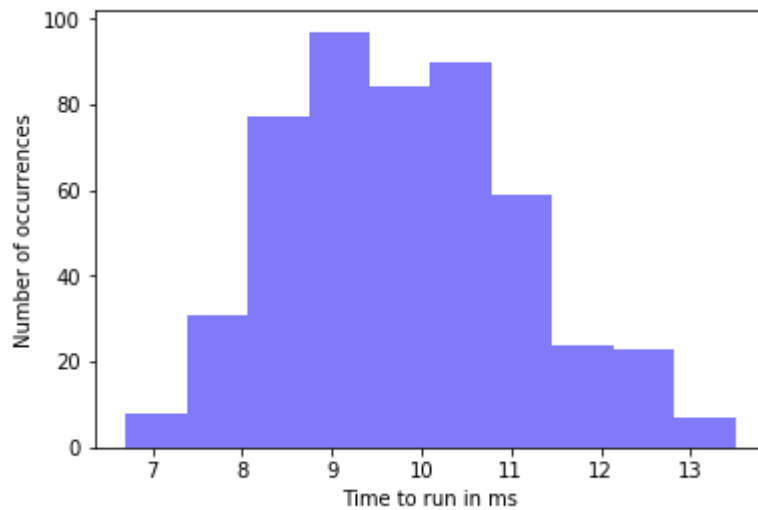
Fig. 2: Histogram of the Monte Carlo method

5.2 Program Code

Java code snippet of the practical implementation:

```java
Cloudlet cloudlet;
List<Double> finishTimeList = new ArrayList<Double>();
for(int i=1; i<=500;i++) {
    // First step: Initialize the CloudSim package. It should be called
    // before creating any entities.
    num_user = 1;   // number of grid users
    calendar = Calendar.getInstance();
    trace_flag = false;  // mean trace events

    // Initialize the CloudSim library
    CloudSim.init(num_user, calendar, trace_flag);

    // Second step: Create Datacenters
    //Datacenters are the resource providers in CloudSim. We need at list one of
them to run a CloudSim simulation
    datacenter0 = createDatacenter("Datacenter_0");
    datacenter1 = createDatacenter("Datacenter_1");

    //Third step: Create Broker
    broker = createBroker();
    brokerId = broker.getId();

    //Fourth step: Create VMs and Cloudlets and send them to broker
    vmlist = createVM(brokerId,20); //creating 20 vms
    cloudletList = createCloudlet(brokerId,40); // creating 40 cloudlets

    Collections.shuffle(cloudletList);

    broker.submitVmList(vmlist);
    broker.submitCloudletList(cloudletList);
```

```
        // Fifth step: Starts the simulation
        CloudSim.startSimulation();

        // Final step: Print results when simulation is over
        newList = broker.getCloudletReceivedList();

        CloudSim.stopSimulation();

        printCloudletList(newList);
        // Get the last Cloudlet
        cloudlet = newList.get(newList.size()-1);
        // Find the time of finishing for the cloudlet. This is simultaneously the
time of completion.
        double finishTime = cloudlet.getFinishTime();
        finishTimeList.add(finishTime);
        Log.printLine("Time of finishing is: " + finishTime);
        //Print the debt of each user to each datacenter
        datacenter0.printDebts();
        datacenter1.printDebts();

        Log.printLine("Round " + i);
        }

Log.printLine(finishTimeList);
Log.printLine("Worst time is: " + Collections.max(finishTimeList));
Log.printLine("Mean: " + mean(finishTimeList));
Log.printLine("Best time is: " + Collections.min(finishTimeList));
```

6   Conclusion and Future Work

This Work provides a brief overview of the Cloud and how it came to be, CloudSim and some of its relevant parts and processes that take place in order to schedule jobs efficiently. CloudSim allows researchers to test their methods and hypothesis in a controlled, highly customizable environment prior to real development so that they can be assured in their results. With a very large number of parameters and an overwhelmingly large search space we implemented the Monte Carlo method for the job scheduling algorithm and got a very noticeable improvement over the non-optimized results even with searching just a fraction of the total search space with a very light weight code.

Future Work includes incorporating metaheuristic search algorithms such as the Genetic Algorithm, Cuckoo Search algorithm or Particle Swarm Optimization instead of the Monte Carlo method. We are looking to achieve even better results by modifying the parameters of the given search functions in order to obtain optimal results. These algorithms should all outperform the current results and give better results with even less computing time. When it comes to running on the large scale every bit of performance scales tremendously and it's essential to be as efficient as possible so that the least amount of time is wasted.

References

[1] Announcing Amazon Elastic Compute Cloud (Amazon EC2) - beta, 2006. [Online]. Available: https://aws.amazon.com/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2---beta/

[2] Navendu Jain, Ishai Menache (2012). *Resource management for cloud computing platforms*. United States Patent No. US9595054B2. [Online]. Available: https://patents.google.com/patent/US9595054B2/en

[3] Antonio Regalado (2011). *"Who Coined 'Cloud Computing'?"*. Technology Review, MIT. [Online]. Available: https://www.technologyreview.com/s/425970/who-coined-cloud-computing/

[4] Eric Jonas , Qifan Pu , Shivaram Venkataraman , Ion Stoica , Benjamin Recht. (2017). *Occupy the cloud: distributed computing for the 99%*, 2017 Symposium on Cloud Computing, p.445-451, September 24-27, 2017, Santa Clara, California . doi: 10.1145/3127479.3128601

[5] Blesson Varghese, Rajkumar Buyya (2018). *Next generation cloud computing: New trends and research directions.* Future Generation Computer Systems, Volume 79, Part 3, February 2018, Pages 849-86. doi: 10.1016/j.future.2017.09.020

[6] R. Buyya, C.S. Yeo, S. Venugopal, J. Briberg, and I. Brandic (2008), *Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility*, Future generation computer systems, Vol. 25, 2009, pp. 599-616. doi: 10.1016/j.future.2008.12.001

[7] Keke Gai, Meikang Qui, Hui Zhao (2018). *Energy-aware task assignment for mobile cyber-enabled applications in heterogeneous cloud computing*. Journal of Parallel and Distributed Computing, Volume 111, January 2018, Pages 126-135. doi: 10.1016/j.jpdc.2017.08.001

[8] *NIST Cloud Computing Reference Architecture*, 500-292, 2011. [Online]. Available: https://bigdatawg.nist.gov/_uploadfiles/M0008_v1_7256814129.pdf

[9] D. Alsmadi, V. Prybutok (2018). *Sharing and storage behavior via cloud computing: Security and privacy in research and practice, Computers in Human Behavior*. Elsevier, Vol. 85, 2018, pp. 218-226. doi: 10.1016/j.chb.2018.04.003

[10] M. H. Ghahramani, M. Zhou and C. T. Hon. (2017). *Toward cloud computing QoS architecture: analysis of cloud systems and cloud services.* IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 1, pp. 6-18, Jan. 2017. doi: 10.1109/JAS.2017.7510313

[11] Ping Li, Jin Li, Zhengan Huang, Chong-Zhi Gao, Wen-Bin Chen, Kai Chen (2018). *Privacy-preserving outsourced classification in cloud computing.* Cluster Computing, March 2018, Volume 21, Issue 1, pp 277–286. doi: 10.1007/s10586-017-0849-9

[12] Sawilowsky, S. S., & Fahoome, G. (2003). *Statistics through Monte Carlo simulation via Fortran*. Rochester Hills, MI: JMASM.

[13] Kleinberg, Jon, Tardos, Éva (2006). *Algorithm Design (2nd ed.)*. Addison-Wesley. p. 464. ISBN 0-321-37291-3.

Dejan BULAJA
Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
dejan.bulaja.16@singimail.rs

Kristina BOZIC
Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
kristina.bozic.16@singimail.rs

Nikola PENEVSKI
Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
nikola.penevski.16@singimail.rs

Converting the Driver Card Data to Readable Format

Claudia Cauca, Roland Meszaros

Abstract

The open borders of the European Union developed road transportation of goods is a thriving and exponential expanding field. It is so important that it has been decided to standardize the regulations applicable in this regard. For the carriers, it is very good news. When getting to know current national regulations, they are also becoming familiar with provisions which are applicable throughout the entire Community. As a result, the drivers' duties with respect to recording their working time are always the same, irrespectively of location. In this project we propose a solution that aims to simplify the process of reading in a human friendly way the data from the driver's card and from the tachograph vehicle unit. The available software on the market which converts the data from the driver's card are really expensive, this being the main reason for developing a software accessible for everyone: authorities, companies and drivers. One of the challenges of the project is to make an easy to use software with a friendly user interface.

1 Introduction

The European Union has established strict rules and legislation that governs the way of driving trucks weighing more than 3.5 tonnes and busses transporting more than 9 persons. Every type of this vehicles has to be equipped with a device that records data concerning the way of their use and whom is driving the vehicles.
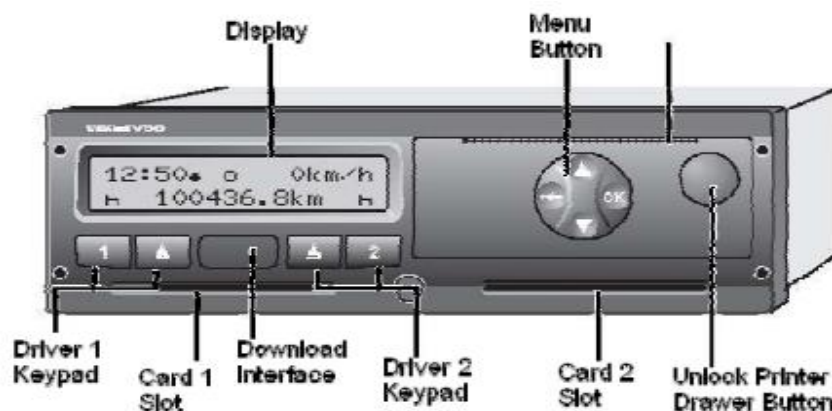
Fig. 1. Front side of a digital tachograph [1]

This type of device is called digital tachograph and is a radio-sized unit coupled with a motion sensor and one or two smart cards [1]. Figure 1 shows a scheme with the frontal view of a digital tachograph.

Digital tachographs aim to improve the old version of this kind of devices which were called analog tachographs. Analog devices used circular paper charts to store the required data. This manner of storing data raised serious discussions and problems regarding the security of stored data. Digital tachographs use smart cards and encryption algorithms to store the driving data. This process makes very difficult the attempt of alter or corrupt the data stored on the cards. Figure 2 shows a diagram of digital tachograph system.
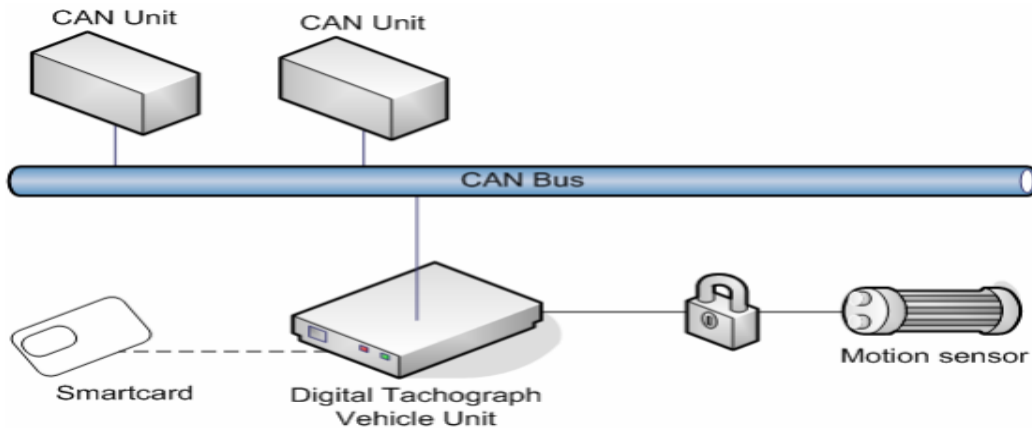
Fig. 2. Digital tachograph system [1]

In 2019 it is planned to replace the digital tachographs with a new generation of devices called smart tachographs which will be able to record professional drivers driving and resting times. Smart tachographs main feature is that will allow automatic recording of start and final location of journeys and will also enable remote access to data. Figure 3 shows a diagram of smart tachograph system [2].
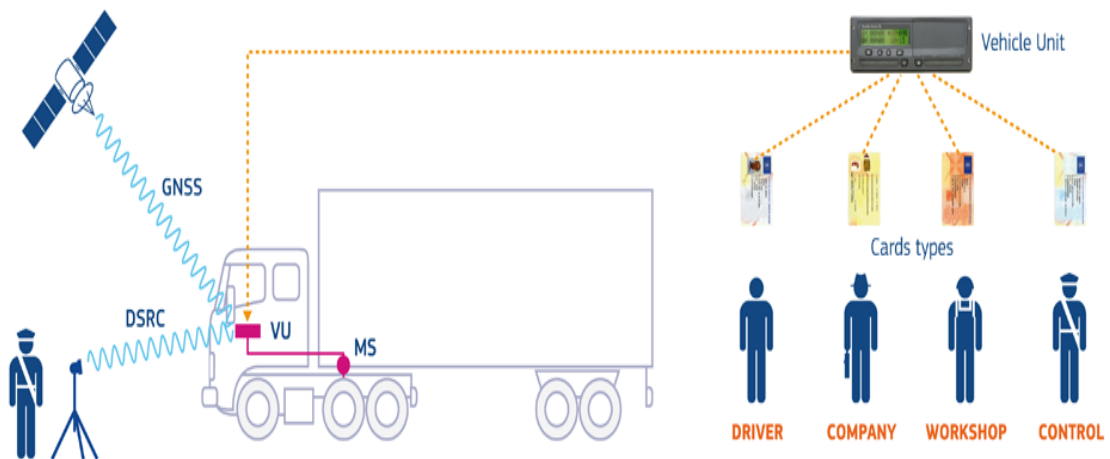
Fig. 3. Smart tachograph system [2]

From the point of view of transport companies one of the main challenge of today use of digital or smart tachoographs remains the integration of downloaded data from a driver card or tachograph

unit. There are several devices available on the market at reasonable prices that allows to download data from driver card and tachograph unit and provides them in a DDD^1 (the file extension containing the data downloaded from the tachograph) file format. This type of file to be read in a human friendly way needs a software that process the data. The price of a software that integrates a DDD file in comparison with the the price of a tool to download the data is very high and not affordable for small transport companies.

This is the main target of this project, mainly to develop a software for integration of DDD files that can be accessible for everyone: authorities, companies and drivers

2   Proposed Solution

A DDD format type is a file that contains the data stored on the internal memory of a digital tachograph or a driver card. The file contains all the information that is related to a vehicle and its driver's related activities and events in a period of time [3]: driving speed, driving time, rest time, driving events, other work time, vehicle malfunctions, co-driver information.
The standard that legislates the structure and the consistency of a DDD file is issued by the Commission of the European Communities and can be found at [4].
Transport companies are legally obliged (according to European legislation enacted in 2006) to download this information on a regular basis and need to provide it on demand to transport authorities during inspections.
There are several devices on the market that provides the download of this data, in the same time there are no open source software integrated in operating systems that can read and display DDD files to a user. This is the main reason for getting this project started.
There are four types of cards that can be inserted in a digital tachograph:
•   driver card – needed by every professional driver in order to drive the vehicle;
•   control card – used by authorities;
•   workshop card – used in workshops for calibration or repair of the;
•   company card – used to download the data from tachograph.
In the first step of the project we target to integrate the DDD file of the driver card. Figure four shows an image of a driver card.

Fig. 4. Example of driver card [5]

All the data in a DDD file is stored in binary values. The target of our work in this stage of the project was to convert the binary values to a human readable format, for which we chose a pdf file format.

First step in the conversion of a DDD file into a pdf file was to convert the data from binary to hexadecimal values which depicts a standard representation for numbers that has 16 as its base.
Next followed the extraction of data according to the file structure specification. Every bit from the file has its own meaning.
After the data was read correctly we had to convert them into integers and strings in order to put the data into pdf file. An example of a function that we used to obtain from binary number the date and time is shown below.

```
public static String date_text(int a)
{
  Date dt3=
  String dateAsText3 = new SimpleDateFormat("yyyy-MM-dd HH:mm:ss").format(new Date (a*1000L));
  return dateAsText3;
}
```

Figure 5 shows the structure of a DDD file with data read from a driver card [4].

- EF Identification
    - CardIssuingMemberState
    - CardNumber
    - CardIssuingAuthorityName
    - CardIssueDate
    - CardValidityBegin
    - CardExpiryDate
    - HolderSurname
    - HolderFirstName
    - CardHolderBirthDate
    - CardHolderPreferredLanguage
- EF Card- Download
    - LastCardDownload
- EF Driving-Licence-Info
    - drivingLicenceIssuingAuthority
    - drivingLicenceIssuingNation
    - drivingLicenceNumber
- EF Events- Data
    - eventType
    - eventBeginTime
    - eventEndTime
    - vehicleRegistrationNation
    - vehicleRegistrationNumber
- EF Faults- Data
    - faultType
    - faultBeginTime
    - faultEndTime
    - vehicleRegistrationNation
    - vehicleRegistrationNumber
- EF Driver-Activity- Data
    - activityPointerOldestDayRecord
    - activityPointerNewestRecord
    - activityDailyRecords

- EF Vehicle – Used
    - vehiclePointerNewestRecords
    - vehicleOdometerBegin
    - vehicleOdometerEnd
    - vehicleFirstUse
    - vehicleLastUse
    - vehicleRegistrationNation
    - vehicleRegistrationNumber
    - vuDataBlockCounter
- EF Place
    - placePointerNewestRecord
    - entryTime
    - entryTypeDailyWorkPeriod
    - dailyWorkPeriodCountry
    - dailyWorkPeriodRegion
    - vehicleOdometerValue
- EF Current-Usage
    - sessionOpenTime
    - vehicleRegistrationNation
    - vehicleRegistrationNumber
- EF Control-Activity-Data
    - controlType
    - controlTime
    - cardType
    - cardIssuingMemberState
    - cardNumber
    - vehicleRegistrationNation
    - vehicleRegistrationNumber
    - controlDownloadPeriodBegin
    - controlDownloadPeriodEnd
- EF Specific-Conditions
    - SpecificConditionRecord
    - entryTime
    - specificConditionType

Fig. 5. Structure of DDD file with data red from driver card [4].

3   Programming tools used in the project

In this project we chose to work with java, eclipse and PDFBox. We chose this environment because it is an open source one and allows advanced file reading and processing. Also it allows transferring the project in a relatively not complicated manner on mobile devices.
Java is a general-purpose computer-programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation.

Figure 6 shows an example of the output of the project.
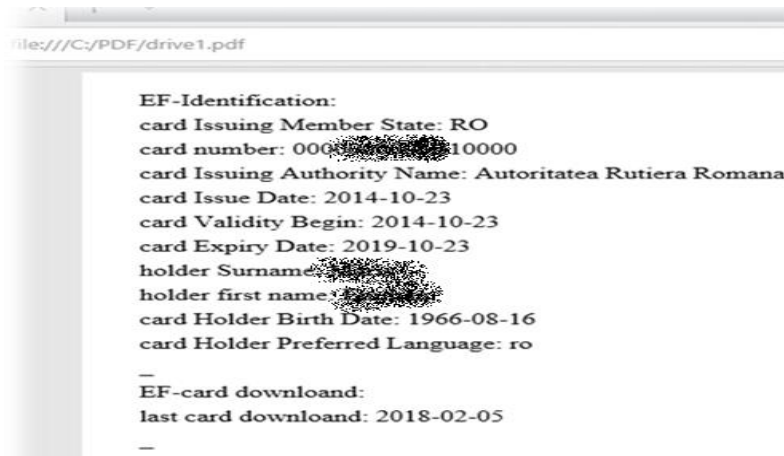
Fig. 6. An extraction from a pdf output of a DDD file

Eclipse is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications, but it may also be used to develop applications in other programming languages via plug-ins.
PDFBox is a Java PDF library which allows working with PDF documents. It allows the creation of new pdf files, the manipulation of existing document and the ability to extract content from a pdf document.

4   Conclusion

Tachographs are design in order to provide an improvement to the safety in transportation sector. Its main function is to record the driving time, breaks, rest periods as well as periods of other work undertaken by a driver of commercial vehicles [6]. With the available data collected by tachographs in a human readable way is expected to discourage infringement of road regulations and to lead drivers to become aware of the risk they expose others and themselves if they do not respect the legislation.

Fig. 7. An example of tachograph software

The project presented in this paper has the purpose to replace other software that read the tachograph's driver card. Also the output of the project can be saved in a pdf file and printed if needed. In the future we intend to extend the project on Android platforms and to design a better, easy to use, User Interface.

References

[1] W. Thomas, *Integrating the Digital Tachograph with Telematics for the new European Standard*, Springer-Verlag, Master of Science Thesis Stockholm, Sweden 2007, available at: http://www.diva-portal.org/smash /diva2:545581/FULLTEXT01.pdf , accessed 03.11.18.

[2] https://dtc.jrc.ec.europa.eu/dtc_smart_tachograph.php, last accessed: 16.11.2018

[3] *What is a Digital Tachograph Container File (DDD File),* available at: https: //uk.support.t elematics.tomtom.com/app/answers/detail/a_id/2091/~/what-is-a-digital-tachograph-container-file-%28ddd-file%29, last accessed: 16.11.2018

[4] *Commission Regulation (EC) No 1360/2002 of 13 June 2002 adapting for the seventh time to technical progress Council Regulation (EEC) No 3821/85 on recording equipment in road transport*, available at: https://eur-lex.europa.eu/eli/reg/2002/1360/oj, last access: 08.11.2018

[5] *Drivers Digital Tacho Card Lost*, available at: https://sevenson.weebly.com/blog /drivers-digital-tacho-card-lost, last accessed: 16.11.2018

[6] D. Borio, E. Cano and G. Baldini, *Speed Consistency in the Smart Tachograph*, Sensors (Basel) 2018 May; 18(5): 1583. Published online 2018 May 16. doi: 10.3390/s18051583

Claudia CAUCA
"Aurel Vlaicu" University of Arad"
Mathematics and Computer Science
Bd. Revoluției 77, 310295 Romania
E-mail: claudia.cauca@yahoo.com

Roland MESZAROȘ
"Aurel Vlaicu" University of Arad"
Mathematics and Computer Science
Bd. Revoluției 77, 310295 Romania
E-mail: rolandmesaros@gmail.com

Constructing a Trustles Medical Data System Using Blockchain Technology

Alexandru Crăciun, Rebecca Monica Pui, Oriana Presacan

Abstract

As technology evolves, the prospect of a world in which all text, audio and video data is easily modifiable by a third party which is capable of accessing it becomes an ever more pressing issue in regards to the safe storage of that said information. This article proposes to take a deeper look at the available technologies used to deal with data modification, by describing and underlining the importance of blockchain technology in this field, while also providing a concrete example to how this specific technology can be implemented, by constructing a program aimed as a medical database, and describing its possible usefulness in regards to safe storage of sensitive information.

1 Introduction

Certifying when a document was created or last changed represents an important matter to be tackled. The main problem faced by researchers and professionals in various fields alike can be the authenticity of the data they are working with, and how that was properly modified. The issue with ever changing information is time stamping the data itself and not the medium.

To combat this ever-present problem, various data encryption algorithms were devised and different programming protocols and methods where implemented, out of which one of the more popular ones available to researchers would be the concept of Blockchain, widely used today in the cryptocurrency industry [1].

As it is, blockchain based applications have a much wider usage in the current digital environment, beyond cryptocurrencies; the applications made possible by the blockchain technology ranging from financial services and even video-games.

1.1. What is Blockchain technology?

The blockchain represents a growing list of records formed out of subunits called blocks, which are linked using a cryptographic key. Each block contains a cryptographic string dependent on the previous block and a timestamp [2]. By its core design the blockchain is made to be resistant to data modification.

On a regular basis, blockchains are typically managed through a peer-to-peer designed network, the communications adhering by convention to a protocol for all inter-node data transfer and for validating any new blocks. The advantage to this type of structure given to the data communicated provides the fact that once data is recorded any other given block dependent on that data cannot

be altered retroactively without alteration of all dependent blocks (see figure 1). Thus, data encrypted into any blockchain structure is secure by design
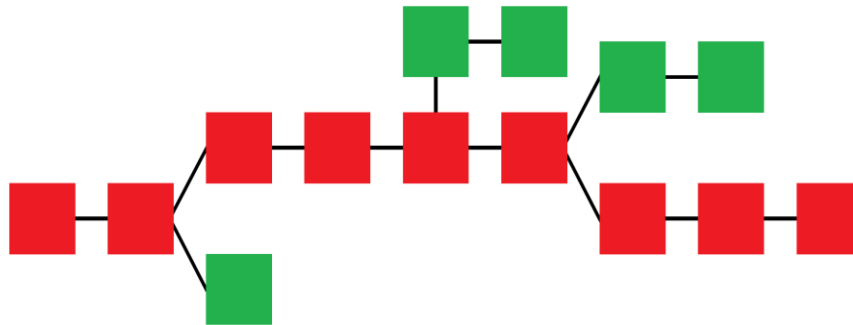

Fig 1. – Blockchain formation; red = main chain; green = orphan blocks

The first recorded usage of blockchain type of technologies goes all the way back to 2008, when a Japanese man going by the alias Satoshi Nakamoto has started his activity ledgering the transactions of the first cryptocurrency, the Bitcoin [3].

It can be affirmed that the invention of the blockchain specifically to be used with Bitcoin made it possible for first digital currency to solve the double-spending [4] problem without the need or implementation of an outside authority in terms of a centralized server.

As it is, Bitcoin operates even to this date in form of both publicly available blockchains and privately-operated ones, with various degrees of usability and reliability, privately owned operations of the sort being described as unreliable by various experts in the field.

2   Blockchain Structure

As it stands, the blockchain is de facto speaking a decentralized public, digital accounting ledger, also collectively distributed, that is used for recording financial exchanges across a network of computers [2], in a way that any involved record cannot be altered retroactively, without the alteration of all subsequent blocks.

The greatest innovation brought forth by the blockchain technology is the possibility to work with a unique digital asset, impossible to reproduce, as each and every transaction or operation takes place. Because of this, the blockchain can be attributed specific title rights as when it is set up accordingly, it can detail the exchange agreement in an unaltered form, as the data it provides is record that compels offer and acceptance embedded into the chain itself, and not through the medium.

2.1 Mechanism

Each individual block of data hold batches of their entire transaction history, data which is encoded into a hash tree [2]. Each block includes a cryptographic string from the prior block in the blockchain, thus the two are always linked via a one-way connection, maintaining the integrity of the entire chain all the way to the genesis block [5].

However, separate blocks can be generated from the same origin block, creating a temporary fork in the larger chain of data. Such blocks which are not selected to be included in the main chain are called orphan blocks (see figure 1).

To ensure a secure transaction history and the integrity of all generated hashes, all blockchains have a genesis specified algorithm for differentiating all possible divergent versions of the chain history so that the block with a higher value can be selected over others. When such situation occurs, the blockchain can form forks in its generation.

2.2. Forks

The Blockchain Fork is defined as following: <a temporary situation that occurs when different blocks are discovered almost simultaneously by different miners> [6]. The fork initially can disrupt the flow of data but is generally speaking quickly resolved when a subsequent bloc is added to the chain, becoming the main chain, the other block remaining in form of an orphan block.
However, forks can also occur not just as an accidental user-dependent fault in generation; they can also appear in form of a change of rules in the software validating the user generated blocks. When this case is applicable, these forks are called hard forks because of their dependency on the software version and algorithms used by the third party hosting the service, rather than an occurrence on the miner˝s side [7].

2.3. Decentralized data

By storing its data all across a peer-to-peer type of network, the blockchain eliminates most of the risks associated with data being held centrally and regulated by a sole entity [2]. Most blockchain based services use a form of ad-hoc message passing protocols and operate on software based on a distributed network.
One of the most important blockchain security models make use of public-key cryptography in combination with a privately generated key, that is dependent on the client side [1].
The public key is de facto an address on the blockchain, whereas the private key acts not unlike a password for the client, giving them access to tradable digital assets or a means to interact the services that the blockchain supports. Because of this double encryption, data stored on the blockchain is considered safe.

3. Medical Databases
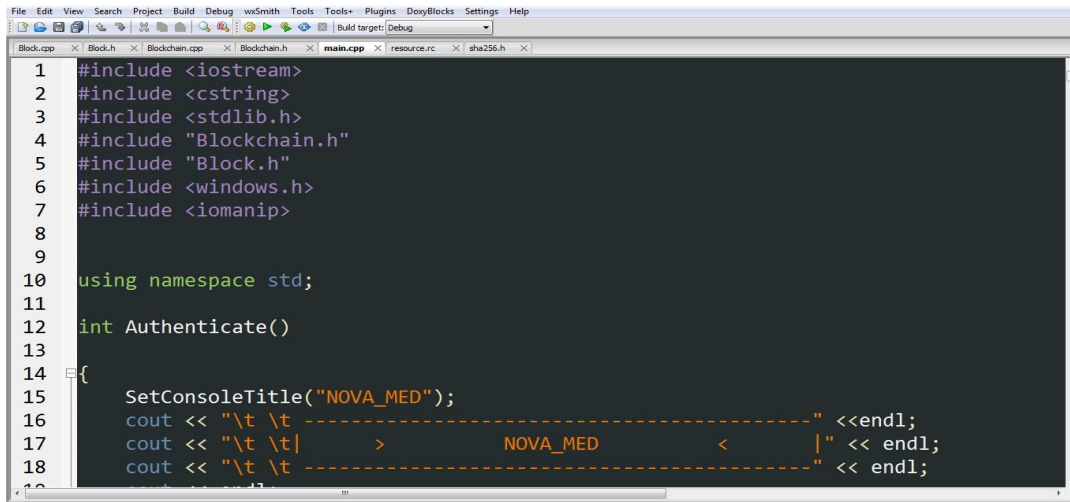
Medical databases, private or publicly kept, are generally speaking operating as bibliographic reference systems, more specifically, they represent a data bank of bibliographic types of records.

As of the latest 40 years, there has been an exponential increase of medical information published and indexed by professionals and researchers alike [8]. Because of this growth, corporations and governments started funding and organizing digital means of storing and indexing all the newly published data, creating collections of references to most published literature, including journal and newspaper articles, conference proceedings, reports, government and legal publications, patents, books available online.

This new storing of data, contrasted to the traditional library catalogue entries, has proven to be a safer and more comprehensive addition to the world of knowledge. A large proportion of the bibliographic records stored in these medical databases describe mostly articles and conference papers, generally containing very rich subject descriptions in the form of keywords, subject classification terms, or abstracts.

4. Implementation

As it stands at the current date, the application we have developed still represents a work in progress and can be subject to further modifications and additions in terms of features before it can be released. The current version is – if anything – a proof of concept.

Fig. 2 – Header and Source files in use.
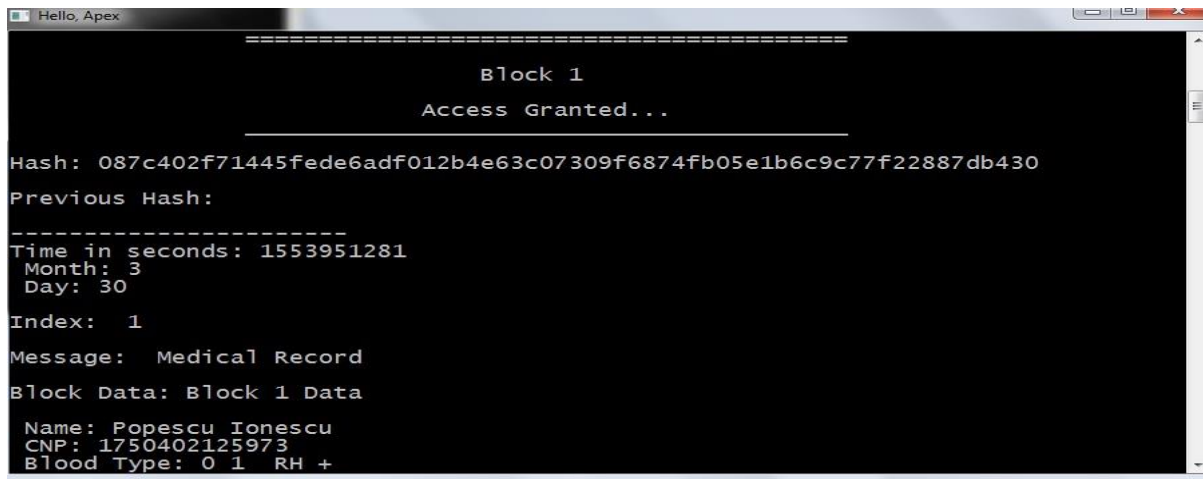
In order to further develop on the idea of blockchain used as more than a cryptocurrency dependent programming abstraction, we intended to further develop on the idea of a database working on functional blockchains – as it was discussed before that it is possible – and thus we have developed a software solution dealing with this idea.

Fig. 3 - The authentication for the Blockchain
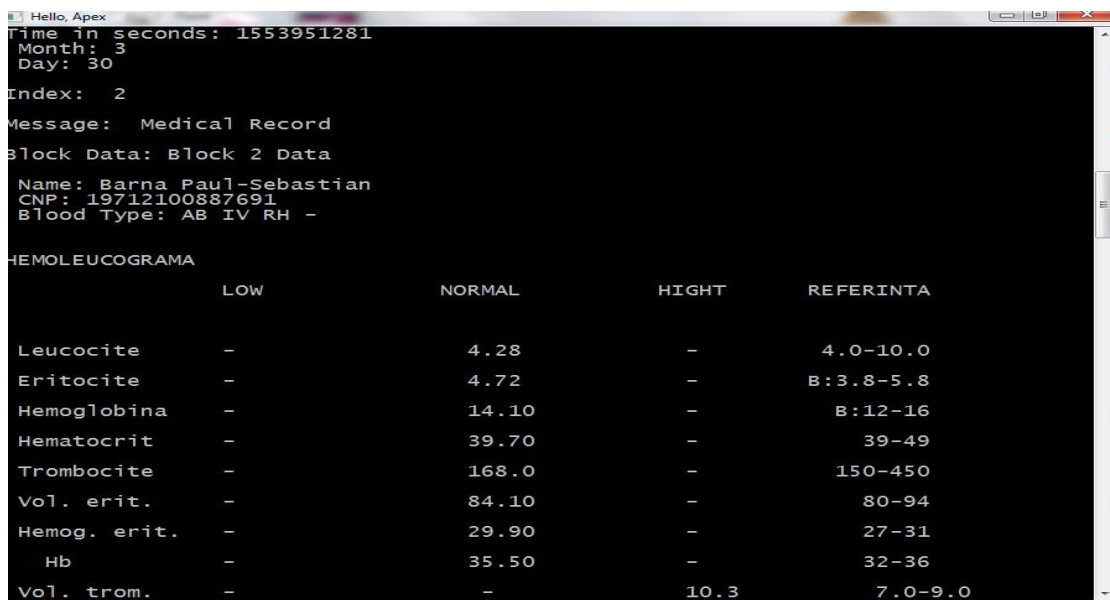
With the use of the C++ programming language the application was developed to the current build and tested. For construction we used multiple header and source files (see figure 2). Our own header files

were used to declare public and private library functions so as to provide easier access to common functions and more accessibility to the code. The „Blockchain.h" source file was used to create the blockchain itself, and the main file was used to put all functions together to create the blockchain.

Fig. 4 - In this figure we can see how it works and how is create a block from the chain

Fig. 5 - Representation of a medical record in blockchain

5. Conclusions

In the current build the designed application proved itself to be a reliable database model, users with limited computer knowledge being capable of using the application in order to generate new and data-secured records.

The Blockchain type of data recording that was implemented provides a secure flow of information which has its metadata coded directly into it, being hardly modifiable, and thus as secure as it can get.

In its current build the software supports up to 9 recorded parameters to be introduced for all patients, while also recording their basic most information such as their identification number, their blood group and their name.

6. Discussion

As it is right now, the application is in its very early stages of development and will further be improved by the means of adding more data fields and more medically significant values to those fields. The graphical interface will also be improved to become more user friendly.

References

[1] S. Haber and W. S. Stornetta, "*How to Time-Stamp a Digital Document*," *Journal of Cryptography,* vol. 3, no. 2, pp. 99-111, 1991.

[2] A. Narayanan, J. Bonneau, E. Felten, A. Miller and S. Goldfeder, *Bitcoin and Cryptocurrency Technologies*, Princeton: Princeton University Press, 2016

[3] The Economist, "*The Great Chain of Being Sure About Things*," 31 10 2015. [Online]. Available: *https://web.archive.org/web/20160703000844/http://www.economist.com/news/briefing/21677228-technology-behind-bitcoin-lets-people-who-do-not-know-or-trust-each-other-build-dependable.* [Accessed 25 03 2019].

[4] N. D. Bhaskar and D. L. K. Chuen, "*Bitcoin Mining Technology*," in Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data, London, Academic Press, 2015, pp. 45-65

[5] U. W. Chohan, "*The Double Spending Problem and Cryptocurrencies*," 19 12 2017. [Online]. Available: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090174.* [Accessed 26 03 2019].

[6] A. Antonopoulos, "Glossary," in *Mastering Bitcoin: Programming the Open Blockchain*, GitHub, 2014, pp. glossary.asciidoc; xvii-xviii.

[7] T. Lee, "*Major glitch in Bitcoin network sparks sell-off; price temporarily falls 23%*," Ars Technica, 12 3 2013. [Online]. Available*: https://arstechnica.com/information-technology/2013/03/major-glitch-in-bitcoin-network-sparks-sell-off-price-temporarily-falls-23/.* [Accessed 28 03 2019].

[8] P. O. Larsen and M. von Ins, "*The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index*", *Scientometrics*, vol. 84, no. 3, pp. 575-603, 2010

Alexandru CRĂCIUN
"Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca
General Medicine
Victor Babeş str. 8, Cluj-Napoca
400000,ROMÂNIA
E-mail: craciunalex98@yahoo.com

Rebecca Monica PUI
"Aurel Vlaicu" University of Arad
Departament of
Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2, Arad
ROMÂNIA
E-mail: nsdproject.vio@gmail.com

Oriana PRESACAN
"Lucian Blaga" University of Sibiu
Department of Computers and Information Technology
Str. Emil Cioran 4 550025, Sibiu
ROMÂNIA
E-mail: orianapresacan@yahoo.com

System for energy consumption minimization for buildings heating

Valentin-Gabriel Crăciun, Paul Adrian Nistor, Mihai Eugen Miuță, Iuliana-Maria Buruiană, Matei Graură

Abstract

The aim of this article is to propose a solution for minimization the energy consumption for buildings heating. With this respect we designed and implemented a system that works like an ―intelligent thermostat". The way that most heating systems work nowadays is by using natural gas to heat water which is then pumped through the installation in order to radiate that heat at the desired places. Although this is generally an efficient solution, it has its drawbacks. Simply installing a heating system fails to take into account the losses generated by thermic energy leaking from the building. This leak is caused by the Second Law of Thermodynamics, which states that energy tends to spread out. This can be mathematically stated by the Clausius theorem. Colloquially, the above formula states that any heat transfer tends to return the corpus that is being heated to its original state. Therefore, the higher an object's temperature is, the more of that heat will be transferred to the environment. As such, an efficient way of reducing that loss is to keep the object's temperature as close as possible to the environment's one for as long as possible. This can be achieved by a self-taught algorithm, which monitors the period of usage any building has and adjust its temperature to minimize losses. Our system uses a learning algorithm. The initial data input is given by the user, which, through a mobile app, gives the system his/her timetable and desired temperature. In time, the system learns the user's habits and makes changes accordingly. The hardware necessary for implementing this solution is cheap and readily available. In contrast to most similar systems, the afore mentioned algorithm requires a small computing unit (like the ever more popular Raspberry Pi) and a number of sensors to be installed both inside and outside of the respective location.

1 Introduction

Every living space needs an optimal temperature to assure human needs. The heating of the buildings involves a high gas or wood consumption. A smart thermostat cannot change the fact that we need natural resources for space heating but by using it, it can greatly reduce their consumption. The motivation of the need for a smart thermostat comes primarily from the desire to save unnecessary heat losses, implicitly natural resources. Natural gas consumption has risen by 96 billion cubic meters (bcm), or 3%, the fastest since 2010. Consumption growth was driven by China (31 bcm), the Middle East (28 bcm) and Europe (26 bcm). Global natural gas production increased by 131 bcm, or 4%, almost double the 10-year average growth rate. The exploitation of natural gases, as we

know, has environmental impacts. The drilling and extraction of natural gas from wells and its transportation in pipelines results in the leakage of methane, primary component of natural gas that is 34 times stronger than CO2 at trapping heat over a 100-year period and 86 times stronger over 20 years. Preliminary studies and field measurements show that methane emissions range from 1 to 9 percent of total life cycle emissions. Natural gas flaring produces CO2, carbon monoxide, sulfur dioxide, nitrogen oxides, and many other compounds, depending on the chemical composition of the natural gas and on how well the natural gas burns in the flare. One recent study found that methane losses must be kept below 3.2 percent for natural gas power plants to have lower life cycle emissions than new coal plants over short time frames of 20 years or fewer. gas development can affect local and regional air quality. Some areas where drilling occurs have experienced increases in concentrations of hazardous air pollutants and two of the six ―criteria pollutants" — particulate matter and ozone plus its precursors — regulated by the EPA because of their harmful effects on health and the environment . Exposure to elevated levels of these air pollutants can lead to adverse health outcomes, including respiratory symptoms, cardiovascular disease, and cancer . One recent study found that residents living less than half a mile from unconventional gas well sites were at greater risk of health effects from air pollution from natural gas development than those living farther from the well sites Also, the exploitation of natural gas destroys the vegetation in the places where geologists explore for natural gas deposits. Well drilling activities produce air pollution and may disturb people, wildlife, and water resources.

Our thermostat attempts to save inefficient heat losses by automating this process. A normal smart thermostat is set by a human according to what he thinks is best for his convenience. Although at first glance this is a good thing, we believe that the cause of the problem is that the majority of people have thermostats manipulated by them to change the temperature. The innovation we bring is the elimination of the human factor. The user has to set just the temperature and the computer knows how to take the decisions to reach the optimal temperature. Mathematical calculations and principles in physics have demonstrated that we can have the desired heat and without a constant heat in the home, considering that the home is not always habited.

Additionally, the heat loss does not affect only the global consumption of natural gases but it means to increase bills unnecessarily. The solution we propose leads to the efficient and rational use of heat taking into account the problems that cause heat loss and the way that the buildings needs to be heated. In this way costs for heating could be significantly reduced.

Another thing that motivate us to continue our project is that the market for thermostats is steadily increasing .The smart thermostat market was valued at USD 689.8 million, in 2018, and is expected to reach a value of USD 2322.2 million by 2024, at a CAGR of 23.1% during the forecast period (2019 - 2024). Lighting and other household appliances consume a considerable amount of energy and presently there is a dire need to use these energy resources efficiently. Therefore, governments are emphasizing on the use of smart devices in homes. The smart thermostat market is highly fragmented due to the government's support and promotion for the uptake of smart devices in residential buildings. Many companies are also entering into the market, making it competitive. Some of the key players in the market include Alphabet Inc. (Nest Labs - Google), Honeywell International Inc., ecobee, Schneider Electric SE, and Lennox International Inc.

We know that the efficient use of natural gas for heating buildings is a small step, but it is also a beginning that can change the course of natural gas exploitation over time. The change comes from us and from our desire to take actions.

2 Technologies

2.1 Thermoregulator

The thermoregulator was realized with "Node.js".Node.js is an open-source, cross-platform JavaScript run-time environment that executes JavaScript code outside of a browser. As an asynchronous event driven JavaScript runtime, Node is designed to build scalable network applications. This is in

contrast to today's more common concurrency model where OS threads are employed. Thread-based networking is relatively inefficient and very difficult to use. Almost no function in Node directly performs I/O, so the process never blocks. Because nothing blocks, scalable systems are very reasonable to develop in Node. Node is similar in design to, and influenced by, systems like Ruby's Event Machine or Python's Twisted. Node takes the event model a bit further. It presents an event loop as a runtime construct instead of as a library. Node simply enters the event loop after executing the input script. Node exits the event loop when there are no more callbacks to perform. This behavior is like browser JavaScript — the event loop is hidden from the user. It communicates with a web server through "Json" packets. The web server has an "UI" and an API.

2.2 Frontend

For realizing the UI the following were used:"React, Redux and Axios" React can be used as a base in the development of single-page or mobile applications, as it's optimal only for its intended use of being the quickest method to fetch rapidly changing data that needs to be recorded. However, fetching data is only the beginning of what happens on a web page, which is why complex React applications usually require the use of additional libraries for state management, routing, and interaction with an API.

React is organized into reusable pieces of code named *components*. They give developers flexibility and allow them to concentrate on every single design piece in isolation from the rest. Furthermore, React apps can take advantage of JSX, which is a syntax extension to JavaScript. After compilation, JSX expressions become regular JavaScript function calls and evaluate to JavaScript objects. It offers the opportunity to treat UI elements and their attached logic at the same time, resulting in a more concise and better organized user interface.

The standard React component is called an *element*. There are a variety of predefined element types, but React's appeal stands in offering an efficient and easy way of creating custom elements. Unlike browser DOM (Document Object Model) elements, React elements are plain objects, their updating in the browser's DOM being handled by the React DOM. Every React app must have a so-called ―root" DOM node, whose component parts are all managed by the afore mentioned React DOM. Also, React elements are immutable. Once an element is created, its children and attributes cannot be changed. Creating the dynamicity inherent to modern UIs requires arranging elements in components and constantly changing their configurations by creating new elements with new attributes.

Redux is a small library with a simple, limited API designed to be a predictable container for application state. It operates in a similar fashion to a reducing function, a functional programming concept.

It is influenced by the functional programming language Elm . Redux bases its utility on the usage of actions. Actions are payloads of information that send information from application to storage. Actions are plain JavaScript objects. Actions must have a type property that indicates the type of action being performed. Types should typically be defined as string constants.

Another useful feature of Redux are reducers. Reducers specify how the application's state changes in response to actions sent to the store. Remember that actions only describe what happened, but don't describe how the application's state changes. They have the advantage of increasing an application's efficiency without importing large libraries or compromising on functionality or design.

The store is an object that remembers all the data. The store has the following responsibilities: Holds application state; Allows access to state; Allows state to be updated; Registers listeners; Handles unregistering of listeners via.

It's important to note that a Redux application must have only one store. If it is desired to split data handling logic across different files, reducer composition is used instead of creating multiple stores. Reducer composition is, essentially, the philosophy that every function must only handle data that is crucial for its functionality and nothing more. This is achieved by having different reducers manage their own part of the global state. In this way, code readability is increased and debugging becomes easier for developers.

Axios is a very popular JavaScript library you can use to perform HTTP requests, that works in both Browser and Node.js platforms.

Using Axios has quite a few advantages over the native Fetch API:
    -has a way to abort a request

-has a way to set a response timeout
-performs automatic JSON data transformation

Axios enjoys compatibility and support across all modern browsers. It is also lightweight and reliable. It is based on HTTP promises. A Promise is a concept for asynchronous operations. Essentially, it represents an object that can be available at any point from now into the future.

It has three states: Pending, Fulfilled (it completed successfully) ,Rejected (it failed), The states of any promise are handled with two methods, then() and catch(). ,then() provides the expected object from the asynchronous call if successful, and catch() will allow handling the error.

2.3 Backend

Services are components that are used by the app. For example, a logging component is a service. Code to configure (or register) services is added to the Startup.ConfigureServices method.

The request handling pipeline is composed as a series of middleware components. For example, a middleware might handle requests for static files or redirect HTTP requests to HTTPS. Each middleware performs asynchronous operations on an Http Context and then either invokes the next middleware in the pipeline or terminates the request. Code to configure the request handling pipeline is added to the Startup.Configure method.

Also, ASP.NET Core has a built-in dependency injection (DI) framework that makes configured services available to an app's classes. One way to get an instance of a service in a class is to create a constructor with a parameter of the required type. The parameter can be the service type or an interface. The DI system provides the service at runtime.

The request handling pipeline is composed as a series of middleware components. Each component performs asynchronous operations on an HttpContext and then either invokes the next middleware in the pipeline or terminates the request.

By convention, a middleware component is added to the pipeline by invoking its Use... extension method in the Startup.Configure method. For example, to enable rendering of static files, call UseStaticFiles.

An ASP.NET Core app builds a host on startup. The host is an object that encapsulates all of the app's resources, such as:
- An HTTP server implementation
- Middleware components
- Logging
- DI
- Configuration

The main reason for including all of the app's interdependent resources in one object is lifetime management: control over app startup and graceful shutdown.

Two hosts are available: the Web Host and the Generic Host. In ASP.NET Core 2.x, the Generic Host is only for non-web scenarios.

The CreateDefaultBuilder method configures a host with commonly used options, such as the following:
- Use Kestrel as the web server and enable IIS integration.
- Load configuration from appsettings.json, appsettings.{Environment Name}.json, environment variables, command line arguments, and other configuration sources.
- Send logging output to the console and debug providers.

3 Thermostat functionality

The thermostat communicates with the web server, from where he is receiving the users preferences concerning the time period and the desired temperature comfortable and suitable for a healthy lifestyle.

Taking into consideration the fact that the house will give up heat according to the exterior environment in order to reach a certain thermic equillibrum, the thermostat will let the house cool-down as long as possible and will allow the heating process to start only when it is strongly needed. Doing this, the house regains the perfect temperature at the perfect time, removing heating-loss.

Our system will test how long it takes for the house to reach a certain temperature in relation with interior-exterior facts. Therefore, the next time you use it in the same atmospheric conditions, the thermostat will know exactly the moment when the heat must be propagated for a perfect timing, memorizing the data.

In short terms, this is how the algorithm works:
– it checks the temperature outside the house and inside the house
– it does periodically tests to ensure if the heat must be spread

And there are 3 cases he takes into account when deciding whether to start or not according to the temperature :

The thermostat is not having the time needed for the house in order to heat memorized. Then, at the moment he starts spreading the heat, he knows that his precision will not be extremely accurate, but the process of memorizing and taking into consideration all the data start. The thermostat will know more and more everyday regarding it's situation in any case.

It does not know for how long will be needed to keep the heat on, but it made an memorized estimation last time. Now, it works according to the previous set of data for this temperature, and at the end of this process it upgrades the data with the new real time estimations for this case. It has all the data memorized so the heat starts according to them.
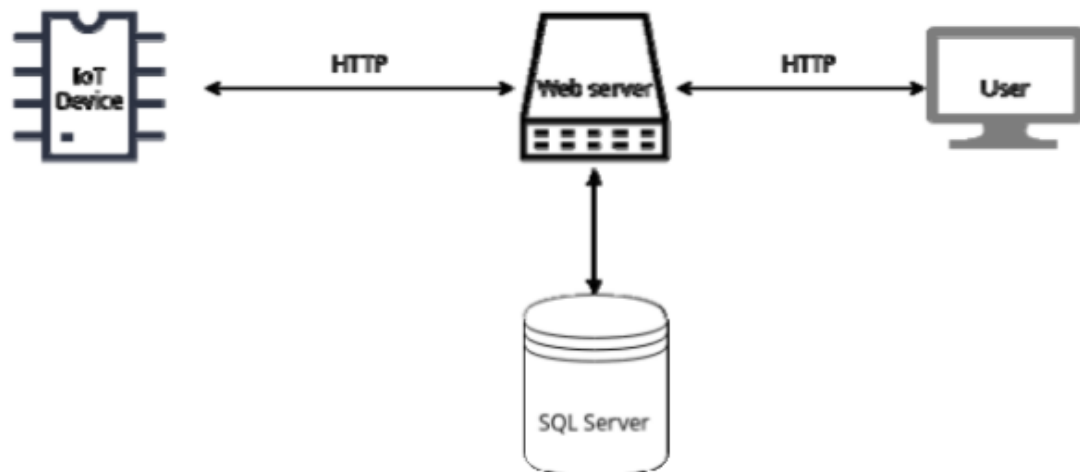
Fig. 1: Architecture

4 Conclusion

In conclusion, we created an experimental project for optimizing energy consumption. For us, to be sure about our work, the next step is to test our presumption and definitivate our work.

References

[1]   https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf
https://www.ucsusa.org/clean-energy/coal-and-other-fossil-fuels/environmental-impacts-of-natural-gas
[2] https://dotnet.microsoft.com/
[3] https://reactjs.org/
[4] https://en.wikipedia.org/wiki/Laws_of_thermodynamics

Valentin-Gabriel CRACIUN
Lucian Blaga University
Faculty of Scrience – Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: Gabriel.craciun@ulbsibiu.ro

Paul Adrian NISTOR
Lucian Blaga University
Faculty of Scrience - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: paul.nistor13@gmail.com

Mihai Eugen MIUTA
Lucian Blaga University
Faculty of Scrience - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu,550012
ROMANIA
E-mail: miuta.mihai@gmail.com

Iuliana-Maria BURUIANA
Lucian Blaga University
Faculty of Scrience - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: buruiana.iuliana.maria@gmail.com

Matei Florin GRAURA
Lucian Blaga University
Faculty of Scrience - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: matei.graura@gmail.com

Bank connect: How to build a fintech app

Gheorghe-Catalin Crișan

Abstract

The goal of this article is to present how to create an application that aims to aggregate user information from the banks where the user has an account. Using this approach, the user can view the balance and transaction history of his accounts in a single application or even make transactions between banks directly from the app. This is also useful when we have to pay invoices, but some providers work only with few banks. Thus, we can pay the invoices directly from the app using other of our bank account. We included also the possibility of helping customers in splitting their outgoings into categories like food, invoices, sports, etc. In this way, the user can check how he spends his money and how he can save some money. The article also presents details regarding the implementation of the application and solutions for handling the session between the client application and banks services. Besides the backend, we'll describe the technologies used for the client app and the motivation for their choice. Our proposed solution is compared with other alternatives for the client app. The comparison revealed that our application has substantial advantages combining all of the features presented above.

1 Introduction

Nowadays, the number of bank transactions made with the card is rising, and fewer people choose to pay cash [1]. Starting from this point, a large number of people have at least two bank accounts that they manage. This means that the person must log into each separate account to perform different actions such as viewing the balance or making a money transfer. This is not very pleasant when the user has to logout to access the other account if both accounts are at the same bank or access another application when the two accounts are at different banks. This problem can be solved by an aggregate banking application like the one presented in this article that allows you to manage multiple bank accounts simultaneously. The way a person manages their income shows how financially savvy he is and also how well he knows how to control his cash flow. In this direction, our proposed application also provides support by guiding users to become aware of the expenses they make. By synchronizing bank accounts, our aggregate application creates a habit through which the user controls his/her financial situation. For example, the user can choose a time of day when he wants to review his account transactions, having the opportunity to organize his transactions by category. In this way, the user has access using a single application to the account information along with the expenditure reports and even the money transfer. Thus, we are able to provide the user with the services he needs along with a spending manager.

Access to user accounts is possible thanks to the European PSD2 Directive [2], which provided the legal basis for the creation of an EU-wide single market for payment services. So the

service contributes to an efficient European payment system, promoting innovation in this area and online payment services, but also introduces the level of security required to protect user data. This service is also recently adopted in Romania, giving the app access to the target market in that country due to the fact that there is currently no direct competition.

2 Problem definition

In Romania, users with multiple bank accounts have at their disposal the mobile applications for each bank, but the user experience is not the best because of the difficulty the user can navigate between accounts. He must log out from the first account to access the second account or to enter another bank if the other account belongs to another bank. Another problem is that the user does not have any automated service to manage his expenses. Existing bank applications do not provide a service allowing the user to access personalized reports of each account so that he can keep track of his expenses. This feature would help users optimize budget, become tidier, and learn to save. More, users could be notified for monthly payments.

The more important drawback of existing applications is that the user must manually introduce each expense into the application. This can be tedious and requires a lot of time, with the possibility of omitting some expenses. The daily effort required for introduction of the expenses in the application can decrease dramatically the user's interest for that application.

The following list presents ones of the most common applications that users can access:

- Monthly Budget Planner & Daily Expense Tracker: allows the users to add their revenue sources that are automatically refreshed every week or month based on what the user chooses. In addition, the user can create budgets for different spending categories. Then, they manually add each expense made to the appropriate expense category.
- Expense Manager: offers similar functionality to the previous application.
- Wallet (www.BudgetBakers.com): offers similar features as our application. The user can synchronize their transactions into the app. One problem could be that this solution does not completely replace the application offered by banks because it does not support money transfer services. Also, there is no integration with Romanian banks.

3 Proposed solution

In order to remove the drawbacks presented in the Section 2 we designed a simple and efficient mobile application intended for Romanian users. Users can add to their application accounts from Romanian banks such as BRD, BCR, ING, Raiffeisen, Banca Transilvania, Alpha Bank, UniCredit Bank, and CEC Bank. Our application provides an integration module which allows users' access to account information, effected transactions, balance and the online payment service. The most important advantage of our solution is the possibility to control all existing accounts regardless of the bank they belong to. We note that our application targets the main banks in Romania and gives users access to all the services present in applications offered by individual banks. Another advantage consists in the support provided for spending management. Users have access to monthly reports for each of their account. Reports are available in two versions:

- Viewing the Monthly Expense Chart: presents the revenue accrued during that month, the monthly expenses organized by category plus the percentage of each category of expenditure

- Viewing the monthly cash flow graph: presents a monthly graph of weekly cash and cash outflows

Adding accounts in the application is secured through the authentication service protocols that are integrated from each bank, such as: registering the application in the bank system, obtaining a public / private key certificate, agreeing the user via a token with a lifetime but also the use of the TLS protocol for the transmission of information. In this way, every third-party application like the application presented in this article must meet a number of validations but also a proper licensing.

Our mobile application solution is developed based on the open source React Native framework. The React Native Framework helps develop native Android and iOS mobile applications by writing JavaScript. Developers write a single code and the application is built for the two mobile platforms, giving the first advantage of less time spending for applications development. Therby, two different applications can be developed for each platform in the same time. React Native is built over the ReactJS framework created by Facebook to develop web applications. Both React Native and ReactJS are based on the same core module called React that can be used to develop mobile applications with React Native and Web ReactJS. This is another advantage of the solution, given that the development team needs to know only the JavaScript language without having separate teams that know Java or Kotlin for Android respectively Objective-C or Swift for iOS. Due to the React core module, we can say that a React Native developer can easily write a ReactJS web application syntax and the patterns used are similar. Therefore React framework is a very good choice in the development of cross-platform applications due to the advantages presented above. For this reason companies like Facebook, Instagram, Skype, Airbnb or Testa use this framework in production.

4   Theoretical aspects

4.1 React Native

4.1.1   JavaScript alternatives

React Native applications are different from other mobile application development alternatives using the JavaScript programming language.

One of the alternatives to develop applications using JavaScript is PWA (Progressive Web Applications) that simulate a native mobile application[7]. The big difference between React Native and PWA is that they run in the browser with the ability to use all the features of a native application even sending notifications to the user's phone [3]. At the bottom of these applications are service-worker scripts that run in the background to enable the native application's functionality, such as server synchronization, data retrieval, caching, etc. The PWA has few disadvantages. One of them is a higher latency due to the way they communicate with the server (they are run by the browser). Other disadvantages are: these applications can not access some mobile device features such as NFC payments, interaction with other applications, smart lock, etc.

The disadvantage of PWA is that it is run by the browser. Because of that, there is a higher latency due to the way they communicate with the server. Also, PWA applications cannot access some mobile device features such as NFC payments, interaction with other applications, smart lock, etc.

Another alternative are web applications that simulate a native application by installing it as a mobile application. The app is downloaded from store and installed as a native application, but all the interaction is done through a web view. For example, Ionic and Cordova are two of the frameworks that allow the development of such applications. Their disadvantage is that the heavy loading is made to a single thread of the application that accesses the web site with the help of the web view.

To avoid all the disadvantages presented before our choice for developing the application is React Native. Using React Native allows the application to be written into JavaScript and is then compiled into native code [4][8]. Thus, the entire UI of the application is native without a web application that is wrapped in a web container as with the other two alternatives listed above.

4.1.2   The Component pattern

React Native uses the Component Pattern [5] on the basis of which each screen of the application is actually a component that contains other components - we can say that "Components are the heart of React". We have two types of components: stateful and stateless. A stateful component has its own state that is internally controlled by the component. In this case, the component declares its initial state at the moment of creation, that is, in the constructor, where it sets the properties it needs for functioning with default values. Therefore, during its existence, a component can update its state, at which point it changes the value of one or more properties of its state. This action can take place when the user inserts a text or triggers a component action, for example.

Stateless components do not have their own status, depending on the parameters another component sends. This means that the component does not internally control its state. Thus, having the two types of components we can make a new classification of their component, presentation component and container component. The presentation components manage how things are displayed without knowing the logic. They only receive the information from a stateful component and display it. Besides information, the presentation components also receive stateful component's callbacks, so by pressing a button for example, the presentation component takes the action and sends it to the stateful component of the behavioural logic. In addition to the presentation, components are the container components that can be seen as parent components. They have one or more presentation components that tell them what information they need to show and what actions they send when they interact with the user. For a better understanding we present in Fig. 1 a code sample for stateful component and in Fig. 2 a code sample for stateless component

```
class StatefulComponent extends Component {
    constructor(props) {
        super(props);
        this.state = { userToken: null }; // initial state
    }

    render() {
        <TextInput
            onChangeText={ (text) => this.setState(prevState => {
                                    ...prevState,
                                    userToken: text
                            })}
            values={this.state.userToken} />
    }
 }
```

Fig. 1 – Code sample for stateful component

```
class StatelessComponent extends Component {
    constructor(props) {
        super(props); // don't have a state, only receive properties via props and display them
    }

    render() {
        <Text> {this.props.userToken} </Text>
        <Button title={this.props.title}
            onPress={this.props.onPress} /> // component don't handle logic, just send back to
```

```
                                //stateful component
        }
    }
```

Fig. 2 – Code sample for stateless component

4.1.3 The Redux store

Redux is a predictable state container for JavaScript applications [6]. It has the role of facilitating the management of the application state. The status of the application can be accessed by components to display and manage the information correctly. In this way, we ensure that all components get the general information they need from the same source.

One important thing to note is that Redux should not be seen as local storage or database for our application. As with the stateful components, the Redux store has an initial state that is set when the mobile application is started. This status may be updated during the application run, and when the application is reopened, the Redux store is initially unaware of the last state that closed the application.

Redux works on the basis of two patterns that update store state[9]. The first is the Action Pattern. Shares are payloads of information that are sent by components to the store. They must have a type that indicates the action type and zero or more parameters that are used to store state.

Reducer Pattern is the other pattern used by Redux. Reducers specify how the application's state changes in response to actions sent to the store. An application may have one or more reducers, each of which has its own state. Thus the Redux store is made up of the state of all reducers from the store. After the action is triggered, the reducer is called. Here, the reducer checks the action type and updates its state based on type and given parameters.

All this is done through components that update the status of the application. Thus, every component who wants to have access to the Redux store must log in to that store and map the property in the store state if he wishes to read information or actions if he wishes to modify information. Code samples for Redux action and Redux reducer are presented in Fig. 3 and Fig. 4.

```
export const setToken = (token) => {
  return {
      type: SET_TOKEN,          //custom action type
      token: token              //data sent to reducer
  }
}
```

Fig. 3 – Code sample for Redux action

```
const initialState = {
  token: null
}

const selectedAccountReducer = (state = initialState, action) => {
      switch (action.type) {
            case SET_TOKEN:              //catch action type
                return {
                    ...state,
                    token: action.token  //read data from action
                }
            default: return state
      }
}
```

Fig. 4 – Code sample for Redux reducer

4.2 Application integration flow

In order to access banks' services, users need to take a few steps:
- Create a developer account on bank website

- Register a new application on bank website and get credentials
- Subscribe to the bank APIs
- Upload client certificate and TLS certificate used for secured access
- Obtain access token based of client certificate for each request
- Consume the bank APIs

Therefore, in order to integrate our application with the services provided by banks, we also must follow the steps outlined above. Because the first three steps are relatively simple, we will not insist on them. After we create the developer account on the bank website we can register a new application. After that, we get a unique Client ID for the app that will be used in every HTTP requests we make to the server. Once the application is registered, we subscribe to the services we want to use in the application. According to the PSD2 directive, each bank must provide three services: access to account information, payment and money transfer and confirmation of funds for transactions.

The next step is the certificates uploading. Here are two certificates based on the Signature Authentication Scheme, one for TLS that is used to grant permission to use encrypted communication, and the other certificate used for signing the token request in accordance with Signature Header RFC. The TLS certificate has a generic HTTPS use and is provided directly by the bank. The sign for the token certificate is generated by the bank for each application is unique. The mobile application must generate a new token for each request based on the certificate, which is added to the request header in the authorization payload.

To generate the token we must execute some steps for each request we made. First, we must calculate the digest of the grant type using SHA-256. After that, we will generate a signature based on the call request endpoint and the HTTP method used. By using the two values obtained above along with the other requests of the header request, ie the current date and time and request UUID, we obtain the access token response.

For now, we can consume bank APIs. All the requests you make have this access token in the header. Thus, we will use it to obtain the user's permission to access his or her bank account through the OAuth protocol, obtaining the bearer's user consent token. Based on the token access for our access to services and user-derived token, we can use APIs to obtain bank account information, transactions, etc.

5   Experimental setup

For the development of the application, the npm package manager provided by Node.js was used to install the necessary packages. The programming environment used is the Visual Code IDE, which helps the developer with a number of extensions that ease the programmer's work. For example, the React Native Tools extension helps debugging and using snippets to generate boilerplate code for React components. Our application is developed on Windows 10, there was no possibility to run and test the application on iOS because of the need for a macOS computer to allow the application to be installed on the emulator offered by Xcode. Therefore, to test the application, the Android emulator provided by Android Studio was used.

To simplify the debugging process, both for the use of breakpoints, the Redux store status view and the changes that take place, as well as for viewing the display properties of the UI components, a react-native-debugger was used, that included in one application the tools offered by React Inspector and Redux DevTools which is presented in Fig. 5.
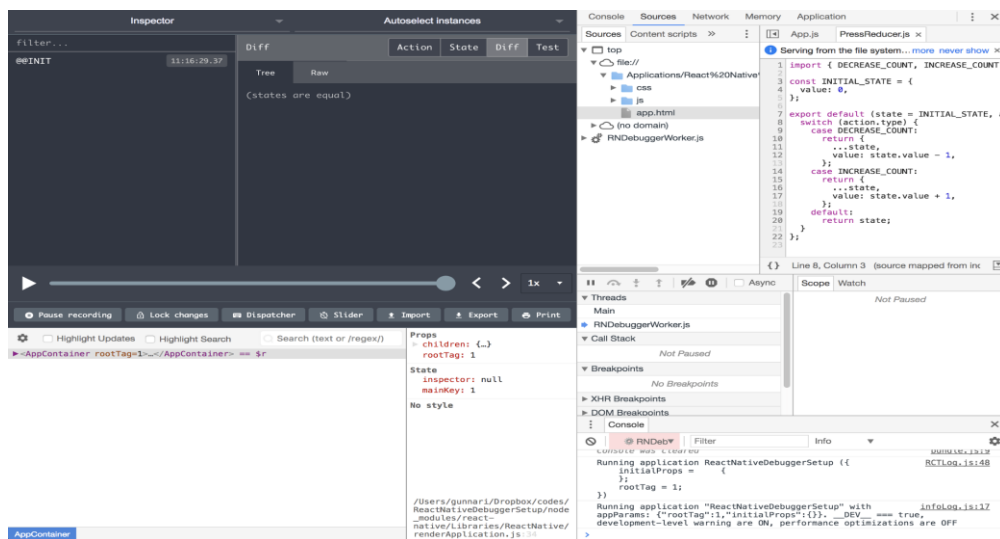
Fig. 5 – React Native debugger

After the part of setting up and configuring the tools needed to implement the application follows the development side. To begin with, configure the Redux store that we access from the react-redux package. The result of this action consists in wrapping the main component of the application into a Provider that receives as a Redux store parameter that we configure with our reducers and actions needed for the store initial state.

To navigate between application screens we use the react-navigation package. It allows us to configure navigators such as stack navigator, which is used to place a screen over another screen and when the current screen is closed we go back to the previous screen tab, which allows us to define tabs to switch from one screen to another or drawer navigator which allow us to create a left side drawer for the application menu for example. React-navigation allows us to define how to transfer the user from one screen to another as in web applications we have URL links that send the user from one web page to another.

Another packet used is async-storage that allows storage of information on the local storage of the mobile phone. This is a persistent storage whose utility is to save user data from one application start to another. Thus, it has access to the history of transfers that have been synchronized from its account, for example. Async storage makes a perfect combination with the Redux store, so the two entities can browse data for the user to have the best experience.

The other two packages used by the application are react-native-vector-icons and react-native-chart-kit that allow us to easily display graphical elements. React-native-vector-icons is used to display icons in the application. It features a number of icons with the design of the Android and iOS operating system. To make a native layout application, the React Native platform API has been used that allows us to display graphical elements depending on the operating system on which the application is installed. Last but not least, the react-native-chart-kit package provides us with a number of graphics components such as line chart, pie chart, bar chart, etc. which allow reports to be displayed to users.

6   Experiments and results

Applying everything presented in previous steps has made it possible to set the objectives of the application, the tools and the packages that the application uses. Next, the application interface will be presented along with the integration of the application into one of the services offered by the bank. The application is compiled from the following screens:

- Home screen: the first screen with which the user interacts. Here he can add a new bank account or select one of the already added bank accounts to see information such as account balance, recent transactions, or syncing his account
- Add bank provider screen: The user selects one of the banks in Romania offering the API service to be redirected to the authentication screen
- Authentication screen: after choosing the bank, the user has to authenticate with the client id and password used for his bank account and is then redirected to the home screen
- Transaction history screen: The user can view their transaction history and filter them
- Categories manager screen: The user can view or update categories to subsequently mark their expenses by being sent to add new category screen
- Add new category screen: On this screen, the user can add a new category or update one of the existing categories
- Monthly report screen: The user has information on earnings, expenses, and balance on the selected month
- Cash flow screen: The user can view monthly cash-flow
- Expenses chart screen: The user can see the percentage of spending on a particular month

The application screens are presented in Fig. 6.

Each bank provides an API specification where it documents the endpoints used by the server to make a request, what each endpoint uses, what are the parameters that awaits them from a request, which is the schema of an object in the request, and which are the responses received along with the corresponding HTTP status code. All these requests are executed via HTTPS and have the form of the request presented in the listing below.

In Fig. 7 is presented a code sample for HTTP consent request. The sent request is based on user's IP address and a date for validity time span.
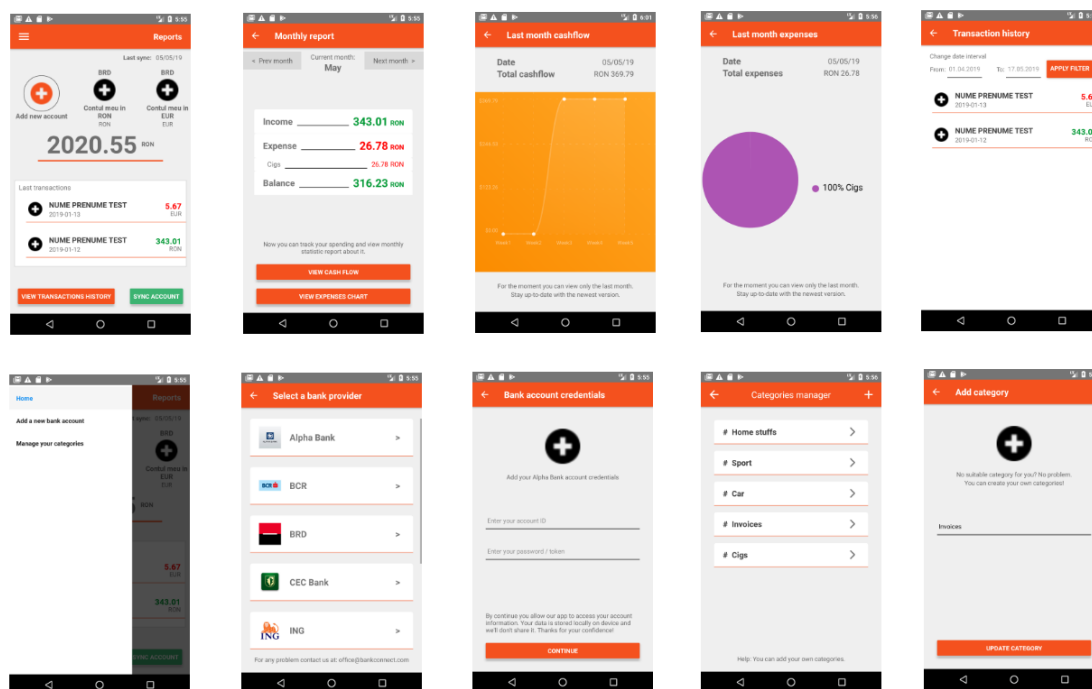

Fig. 6 – Application screens

```
const getConsent = (psuIpAddress, validUntil) => {
    return fetch(API_URL + "consents",
        {
            "method": "POST",
            "mode": "cors",
            "headers": {
                "accept": "application/json",
                "content-type": "application/json",
                "psu-id": CLIENT_ID,
                "psu-ip-address": psuIpAddress,
                "x-request-id": uuid.v1()
            },
            "body": JSON.stringify({
                access: { allPsd2: "allAccounts" },
                recurringIndicator: true,
                validUntil: validUntil,
                frequencyPerDay: 4,
                combinedServiceIndicator: false
                })
        })
        .catch(err => throw(err))
        .then(res => res.json())
        .then(parsedRes => { return parsedRes }
) }
```

Fig. 7 – Code sample for consent request

7 Conclusions

The aim of this article is to present how to build a fintech mobile application for the target market of Romania. In the first part, we presented the benefits of a native hybrid mobile approach for the two Android and iOS platforms. In the second part, we explained all the steps needed to develop a mobile application and how to integrate such an application with the services provided by banks.

Based on our solution, we can be competitive in the fintech area with applications from other countries. We consider our application has an advantage in the target market of Romania. Therefore, is possible to realize an autochthonous application to support the people with accounts at the Romanian banks. In this way, they can manage the accounts easier.

The application at the level that is achieved so far is functional and allows us to achieve the main objectives that we have considered in its design. The application can be developed by adding extensions to improve its functionality. The main extensions we would like to implement in the future could be to integrate similar services such as Revolut and add extra functionality to users' reports. In this way, they become more financially disciplined and can better organize their expenses. In addition, we also want to add a billing payment feature to remind users of the billing period and the ability to quickly pay with QuickPay.

References

[1] STARTUP CAFE, *Crestere de 75% a numarului de plati digitale*, https://www.startupcafe.ro/stiri-ecommerce-21665571-crestere-75-numarului-plati-digitale-anul-trecut-vacantele-taxele-donatiile-generat-cele-mai-mari-plati-online-cardul.htm (accessed 05.05.2019)

[2] EUR-Lex, *Revised rules for payment services in the EU*, https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=CELEX:32015L2366

[3] Hackernoon, *Progressive Web Apps – The Future of Mobile Web App Development*, https://hackernoon.com/progressive-web-apps-the-future-of-mobile-web-app-development-f29257b0dea2 (accessed 05.05.2019)

[4] Medium, *React Native: What is it? and, Why is it used?*, https://medium.com/@thinkwik/react-native-what-is-it-and-why-is-it-used-b132c3581df (accessed 05.05.2019)

[5] Medium, *React component patterns*, https://medium.com/teamsubchannel/react-component-patterns-e7fb75be7bb0 (accessed 05.05.2019)

[6] Redux, *Official documentation*, https://devdocs.io/redux/ (accessed 05.05.2019)

[7] Tal Ater, *Building Progressive Web Apps*, O'Reilly Media, Inc., 2017.

[8] Bonnie Eisenman, *Learning React Native*, O'Reilly Media, Inc., 2015.

[9] Daniel Bugl, *Learning Redux*, O'Reilly Media, Inc., 2017.

Gheorghe-Catalin CRISAN
Lucian Blaga University of Sibiu
Faculty of Science
5-7, Doctor Ion Ratiu Street, Sibiu
ROMANIA
E-mail: crisan.gheorghecatalin@gmail.com

Face recognition framework
based on features template matching

Aleksa Cuk, Branivoj Miljkovic, Milos Todorovic, Aleksandar Ivanovic

Abstract

Facial recognition represents an automatic way of identifying or profiling a person using their digitalized portrait. It consists of many algorithms used to identify the person, but in its core it still consists of the comparison of the offered sample picture obtained through a camera and their predefined picture in the database. It finds application in systems where security is crucial, concretely in two ways: recognition and verification. When considering recognition, it is based on obtaining the facial characteristics of a person using the acquired sample picture. As for the verification the gathered parameters are compared with the already existing picture of the same user in the database. Both of these steps are needed to ensure that the individual who is requesting access is in fact someone with allowed permission. Facial recognition might be less precise compared to iris or fingerprint scanning, nonetheless it is accepted for its simple and convenient way of use.

1 Introduction

Some of the typical issues we face when keeping track of attendance are reliability and efficiency. It is more convenient to track attendance of students on classes using facial recognition. This lead to the development of our program, where the primary functions are identifying and registering the attendants. By automating the whole process, we make it substantially better and more efficient.

The first requirement that needs to be fulfilled is that the students need to be registered in the database. In the case that they are not, they should apply to be enrolled in the database.

The obligatory hardware for our project consists of a camera used for the physical part of the registration, that contains the necessary software. There are no mandatory physical prerequisites for the camera. As for the software, there is the communication between the application and the database, which consists of the registered users. When beginning the authorization, that is, when the camera is approached, the user is prompt with a message that informs them to remove any objects that can obscure the camera (e.g. glasses - in the case that they were originally registered without them) and to frontally face the camera for the sake of precise recognition. After the recognition, the user gets a message from the device about the successful registration.

Finally the user will be informed if they were registered or not.
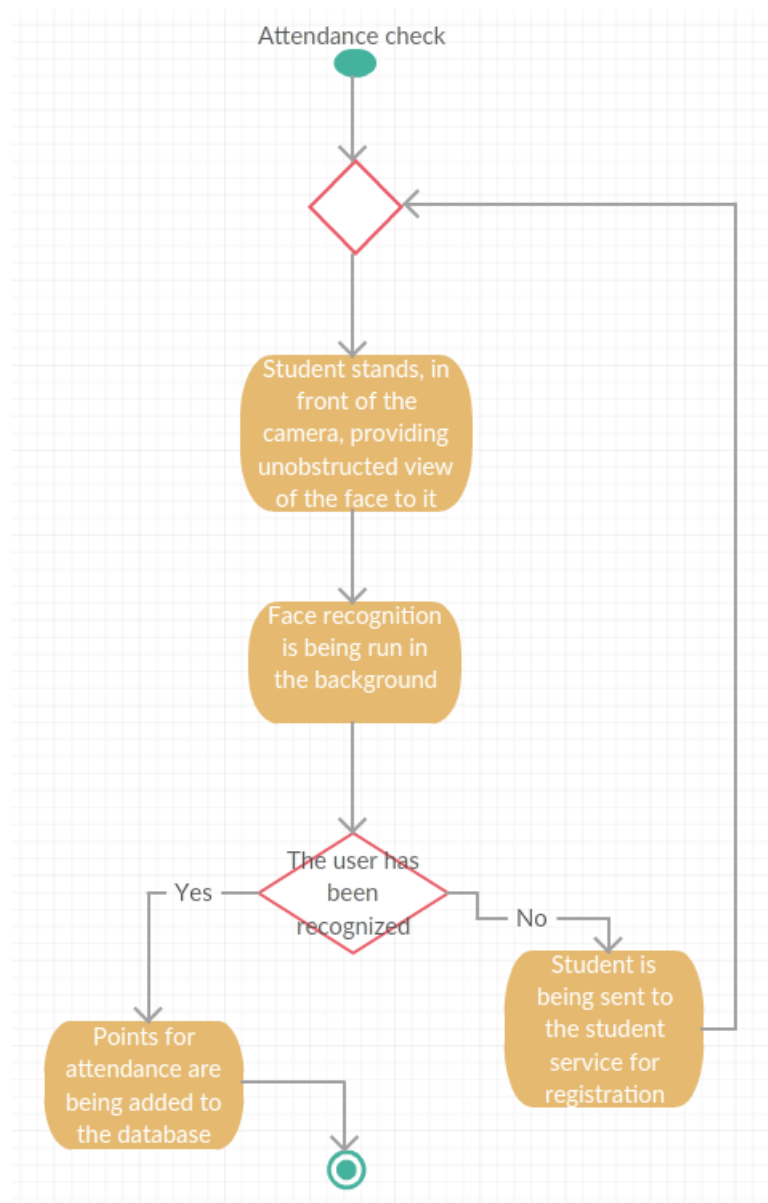
Fig.1 : Program structure

2 Facial recognition

Face recognition (FR) is one of the techniques of biometric portfolio, that is able to verify and identify a person by analyzing and comparing the user's facial features. In recent times, face recognition has been mostly applied in the field of computer security, but it also found application in other fields. For example it is used in payment systems, for the purpose of account verification, as well as game development and modeling. The methods of facial recognition that are in broad use are adaptive regional blend and generalized matching face detection methods.[7]
The staple point of every face recognition based system is the use of sixty eight nodal points of human face that uniformly and uniquely identify it. Values are mathematically mapped and the data is stored as a faceprint.

3 Deep learning

Deep learning is on of many branches of machine learning, which found its appliance in artificial intelligence, where human like way of thinking is applied to algorithms which preform calculations and gain knowledge from enormous datasets which they can apply in the future.[1] Deep learning makes a part of facial recognition purely based on its neural networks that have many layers that offer different forms of learning. Deep learning algorithms obtain information like humans, that is, they learn from experience. They perform best when the sheer amount of data is exceptionally big (e.g. computer vision, processing of speech etc.).

In general, deep neural networks became a dominant technique in machine learning, and their appliance in facial recognition is yet to reach its limits. [3]

4 App description

Facial recognition can be classified as a series of several connected problems. The first step in finding all the faces in the obtained picture. After that it analyzes each face and gathers information from it, no matter if person might be facing the camera from a weird angle or the lighting might be bad. In the third step the program should pick out unique facial features for each person so it is able to differentiate them apart. And finally it will compare the current user's picture with the pictures of all of the registered users, to determine if they are registered or not. All of these steps are defined individually since computers are not capable of high-level generalization, so we need to pass the result from each step to the other.

4.1 Finding faces

For identifying faces in a picture we use a histogram of oriented gradients (HOG). To enable HOG to work, we first have to make the picture grayscale, that is, to ensure normalized color and gamma values. This is necessary so that we are able to identify how dark are the pixels compared to pixels that are surrounding it. After the evaluation, the pixel will be replaced with an arrow that will indicate in which way the image is getting darker. We will repeat this process for each pixel in the image, until all of the them are replaced by arrows. This will provide us with the gradient of the picture. Since having the gradient for each pixel in our image provides us with too much detail, we will transform the picture into a square of 16x16 pixels and each square will count the number of gradients pointing in the major directions.[6]

Regarding the percentage of successful recognition, we achieve 99,38%. This is accomplished by comparing the person that is in the given moment in the objective of the camera with the already registered users in the database.

4.2 Posing and projecting faces

After obtaining the faces, we face the problem of dealing with the positioning of the users head and the direction in which they are facing the camera. The solution for this comes from the 68 parameters which can be uniquely identified for each and every person. Concretely, they are based on top of the chin, outside edge of each eye and the inner edge of each eyebrow. We gather these 68 parameters by using the facial landmark estimation. Then they will be passed off to our machine learning algorithm so we train it to find these 68 parameters on any face.

After we isolated the faces in our image, we have to deal with a problem with direction of the face. The basic idea is, we will come up with 68 points that exist on every face, top of the chin, outside edge of each eye, the inner edge of each eyebrow. We will obtain this 68 point using the face landmark estimation. After that we will train machine learning algorithm to be able to fine these 68 specific points on any face. [5]
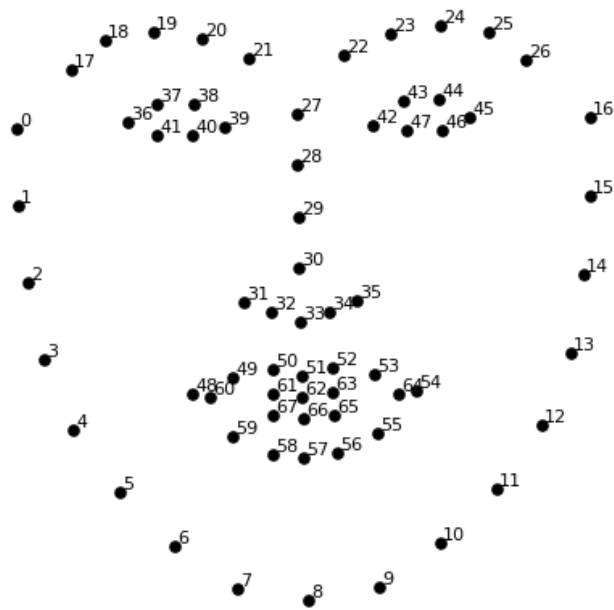
Fig.2 : 68 face landmarks

4.4 Encoding faces

The next obstacle we face in our path is the problem of telling faces apart. This is managed by directly comparing the parameters of the current user, that we obtained from the second step, with the parameters of already registered users. But these parameters are not enough, so we would need to turn to more precise measurements. That is we will take the 68 parameters and turn them to the 128 measurements for both the current user and the people in the database and by comparing these we will be able to precisely tell each person apart.

Fig.3 : Picture of two persons: one register and one that is not

Fig.4 : Person before and after registered in database

5 Implementation

The two core elements of our application for tracking attendants on lectures are Python and OpenCV library.

5.1 Python

Python is a high-level programming language released in 1991 by Guido van Rossum. It is a general purpose programming language which supports object-oriented, functional, procedural and imperative programming.[2]
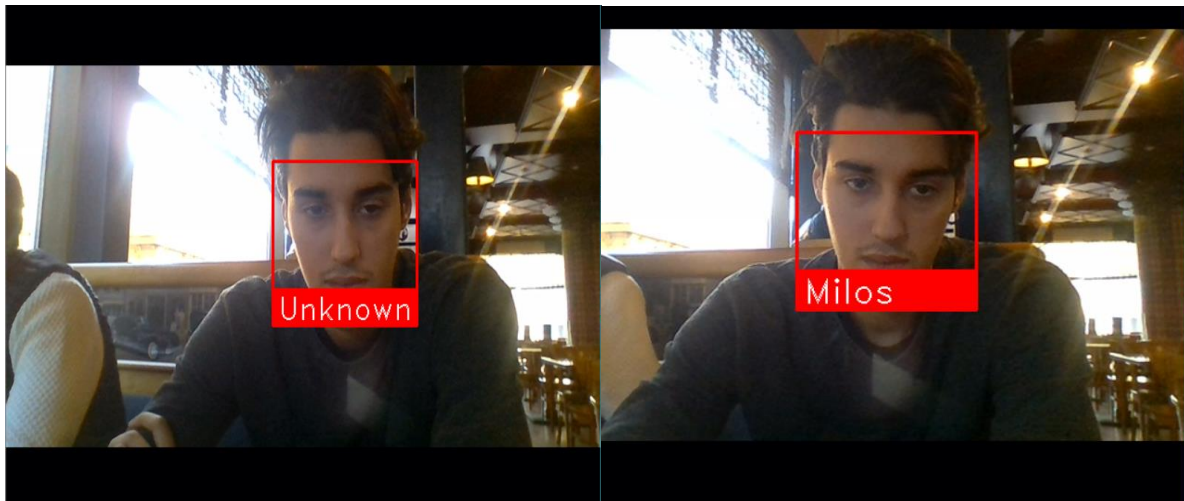
5.2 OpenCV

Open Source Computer Vision Library (OpenCV) is an open source computer vision and machine learning library that was natively written in C++ and that works hand to hand with STL containers. It is mostly used in Python but it can be also used in JAVA, MathLab and it is supported by Windows, Linux , macOS and Android operating systems.

The number of optimized algorithms is astonishingly high, more than 2500, the library also includes comprehensive set, of both of classical as well as state of the art computer vision and machine learning algorithms which can be used to detect faces, classify human action, identify and track moving objects.[4]

6 Application in the real world

The usage of intelligent software is seeing an expansion in use in recent years and it is slowly becoming a part of our daily routine. Its appliance in the process of attendance tracking would improve both efficiency and reliability.

Applications of face recognition:

- Prevent retail crime : face recognition is currently being used as an attempt to lower crime related to retail stores. Photographs of the individuals who engage in crime related activities can be matched with their mug shots to determine if the individual is a customer or a potential shop-lifter. These types of systems are lowering crime activity and making citizens feel safer.

- Unlock phones : the most common and well known usage of face recognition these days is in the security system of smart phones, where face recognition is a powerful way to protect sensitive data from malicious users.

- Smarter advertising : many companies in the world are considering face recognition for the use of advertisements, in that scenario the face recognition will be used to identify the gender or age which will be used to target potentially more suiting customers.

- Find missing persons : maybe one of the most philanthropical application of face recognition would be in the reduction of human trafficking where the missing person's photograph can be compared with the ones in the database.

- Protect law enforcement : by being able to identify people from a distance, using face recognition, police officer will be alerted if the individual has a criminal record and therefore can approach the situation adequately.

- Identify people in social media : face recognition is already being applied on social media (e.g. Facebook) to immediately identify the persons in the photo and provide smarter tagging which can help them later on find their own photos.

- Diagnose diseases : we can diagnose if a person is in the beginning phase of a certain disease which manifest through their face. An example would be a rare disease DiGeorge syndrome.

- Protect school from threats : by labeling expelled students, hostile parents or potentially dangerous individuals which would be added with a connection to the police database, we can prevent them from entering the schools perimeter.

- Casinos : security in a casino can be informed about known frauds and cheats, if they are recognized by the cameras which are connected to a database in their system.

- Control access to sensitive areas : it can be used as one of the steps in security protocols of high level priority government or military facilities to provide access to sensitive areas. Where face recognition can be used in combination with keycard, PIN codes, etc.

7 Conclusion and future work

The goal of our project is to modernize the way of recording attendants at lectures. There are many ways that this is done today, from using personalized cards to writing down signatures, but we came to the conclusion that face recognition can be more precise and practical than any currently known conventional method. We found that using facial recognition is quite challenging and we tackled many problems throughout the development. The first major decision that we had to make as a team was do we use a face recognition algorithm or to opt for face encodings. Through detailed analysis we came to the conclusion that the initial suggestion is more suitable for our application. The next decision that we faced was to either implement the pictures directly into our code using direct pathing or to implement a database. Even database implementation made the program more complex and made our development and testing extensively longer and more challenging, we found that it is better for larger number of users, safer and more reliable than supplying the code with the jpgs. There was a lot of room left for improvement and additional testing. That includes testing the consistency of the program when the number of users becomes exceptionally large.

References

[1] Milosavljević M. ,*Artificial Intelligence*, Singidunum University, Belgrade, Serbia, 2015.
[2] Chollet, F. , *Deep Learning with Python*, 2017.
[3] Shanmugamani, R. , *Deep Learning for Computer Vision*, 2018.
[4] Rosebrock, A. , *Practical Python and OpenCV*, 2014.
[5] Rosebrock, A. , *Deep Learning for Computer Vision*, 2017.
[6] Emami, S., & Suciu, V. P. , Facial recognition using OpenCV. *Journal of Mobile, Embedded and Distributed Systems*, 2012.
[7] Kar, N., Debbarma, M. K., Saha, A., & Pal, D. R. , Study of implementing automated attendance system using face recognition technique. *International Journal of computer and communication engineering ,* 2012.

Aleksa ĆUK
Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
Serbia

Branivoj MILJKOVIĆ
Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
Serbia
E-mail: branivoj.miljkovic@gmail.com

Aleksandar IVANOVIĆ
Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
Serbia

Miloš TODOROVIĆ
Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
Serbia

Cross-platform mobile application for remote-controlling a PC via the Internet coordinated by a RESTful API

Elian Doran

Abstract

With the advent of the Internet of Things (IoT), devices have become increasingly inter-connected but PCs have remained remarkably stagnant. This paper explores increasing the grade of inter-connectivity between mobile phones and PCs/laptops by allowing a mobile user to control their PCs remotely via the Internet or by using a local wireless network. Multiple facets are presented and integrated into a unified application, including: multimedia controls, system information and control, virtual mouse and a proposal for Wake-On-LAN via the Internet.

1 Introduction

It seems that the current market for PC remote control application is quite mature, judging by the number of search results and ratings for "PC remote" in Google's Play Store. Nevertheless, there are two different aspects upon which it could be improved:

1. While there is a high number of features that these remote control applications provide, it usually is the case that a user has to use multiple remote control application to obtain the desired level of control. For example, during a technical presentation one would use *Microsoft PowerPoint: Slideshows and Presentations*[1] to switch between slides and at the same time have to use *Remote Mouse*[2] to present an application. Having to use multiple tools to reach a goal can be difficult at times, especially since these tools come from the same family and could have been integrated into a singular application.

2. Most of the remote control applications have proprietary source code or are simply closed-source. For more technical users, this can prove to be a problem since there is no way for the user to determine that the application they are using correctly handles sensitive information (such as keystrokes, screen imagery, etc.). Without the possibility of investigating exactly what an application does, it can later be revealed that an application used insecure protocols or have backdoor functionalities.

This paper proposes a unified approach towards controlling a personal computer remotely using mobile technologies. One of the most important aspects of this project, however, would be the fact that it is an open-source software and as such it is completely transparent in its implementation, especially useful when dealing with sensitive topics such as remotely controlling a device.

---

[1] https://play.google.com/store/apps/details?id=com.microsoft.office.powerpoint
[2] https://play.google.com/store/apps/details?id=com.hungrybolo.remotemouseandroid

1.1 Structure of the article

1. Introduction	The motivation behind this paper, as well as a summary of the paper.
2. Functional Overview	A presentation of all the features that are currently implemented, separated by UI screens. For each component presented, suggestions of further improvement are also mentioned.
3. Technical Overview	A description of the underlying technical stack comprising the remote control system.
4. Conclusions	Advantages/disadvantages of the project, planned features and final words.

2 Functional Overview

2.1 Home screen

The navigation of the application is tab-based, most screens of the application are formed of the following components:

1. The *app bar* is usually composed of the page title, but in the case of the home screen it displays the name of the PC as it was defined in the user settings. In the case of the home screen, the app bar is expanded in height when first shown and gradually decreases its size as the users scrolls to reveal more content.

2. The *navigation bar* displays icons to the other tabs of the application. Each tab has its own navigation history and in order to avoid usability issues swiping between tabs is not allowed, the only way to switch between tabs is to use this navigation bar.

The home screen offers a series of widgets which allow modular control over its target computer. At the moment the following widgets are implemented:

1. *Volume and brightness control* – displays two sliders the user can use to change the global system volume (0-100% volume) and the brightness of all displays (if the displays support brightness changes, such as laptop displays).

2. *Media player* – connects to a media player (in this particular implementation VLC) and displays the active media title (or simply "No media" if none was found) as well as buttons to navigate the playlist (go back, go next) with pause/resume.
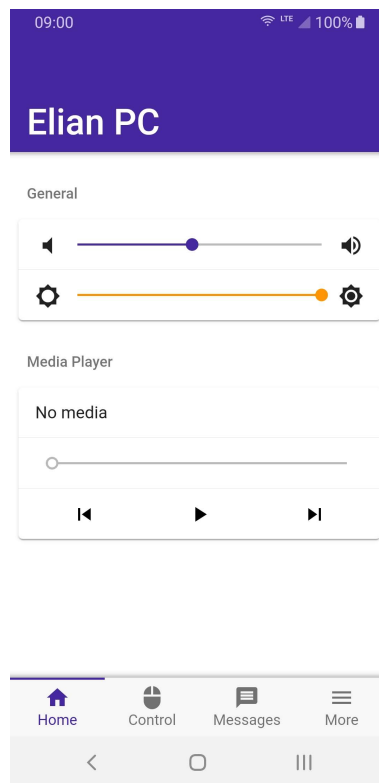
Figure 1: Screenshot of the screen that is first shown when the user accesses the application.

2.2 The control screen

The control screen represents the most important feature of the application, since it allows greater control of the remote PC simply by emulating the hardware mouse and keyboard.

The screen is split vertically into two components:

- An empty space which acts as a software trackpad. Dragging a finger across this area causes the mouse pointer on the PC to slide a certain distance in the same direction. Tapping on the touchpad would generate a mouse click, whereas long pressing would generate a right click.

- A virtual keyboard that is fixed to the screen, unlike standard pop-out Android or iOS keyboards. Upon pressing a key on this keyboard, it would be sent directly to the remote PC.

In order to improve the user experience, instead of displaying a blank space for the touchpad, an instruction text is provided to let the user understand that the area can be used as a touchpad. In the future, it is planned that the rather blank touchpad area will be turned into a live feed the remote PC screen. In order for this feature to be more useful, the live screenshot functionality would be zoomed-in a certain percentage (user-configurable) and centered to the current mouse position.
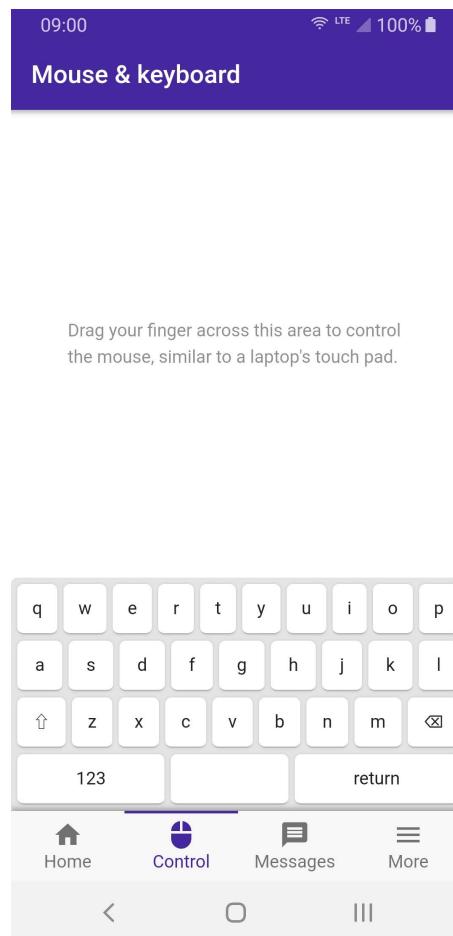
Figure 2: Screenshot of the control tab.

2.3 The messages screen

This screen allows sending messages back and forth between the mobile application and the remote PC. The following message formats are supported so far:

1. Plain text – the message is displayed as-is, with word wrapping;

2. URLs – any URL in a plain text message is automatically converted to a clickable URL;

3. Images – a smaller version of the image being sent is displayed in the message area and it can be clicked to reveal the full image.

4. Files – a small icon representative of the file type, the name of the file and the its size are displayed with the download starting upon clicking.

On the computer side, messages sent from the mobile application are displayed in a notification but they can also be accessed via a dedicated web application which offers a very similar user interface to the mobile version.

Sending messages from a computer would generate a notification on the mobile application which upon opening displays this particular screen.

This particular functionality is very similar to what *Pushbullet* offers, but in this case it is offered as a bundled functionality of the remote control application, rather than a stand-alone product.
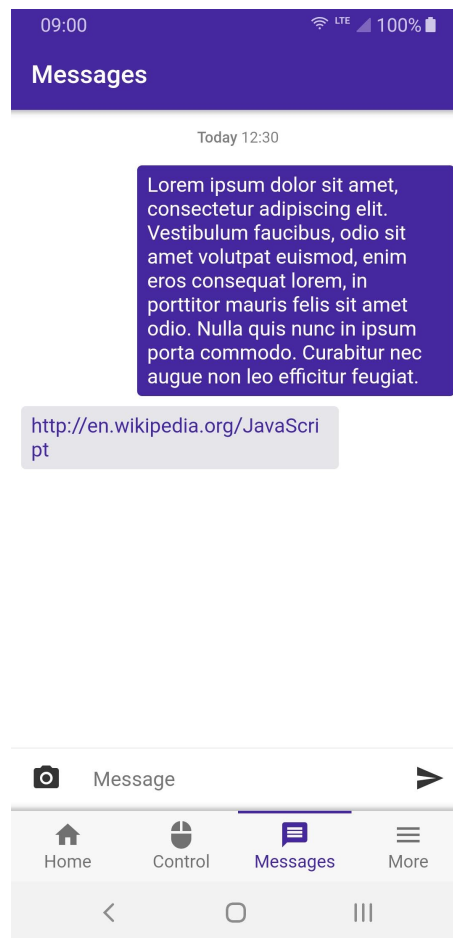
Figure 3: Screenshot of the messages tab.

2.4   The additional options screen

This particular screen was designed as a catch-all for all other functionalities which did not fit in the *Home*, *Control* or *Messages* tabs.

   At this time, the following options are provided:

1. A system information pane, which displays the current CPU, RAM and disk usage in both graphical and text representations.

2. A navigation list for tools that are not as often used as the ones presented above. More tools will be added in the future, but the following are currently supported:

    (a) *File Explorer* – allows the mobile application to remote browse the file system of the remote PC with standard or administrative rights. Transfer operations (such as downloading or uploading files to/between the mobile and remote devices) are supported, as well as remote operations (remote delete, create directory, etc.).

    (b) *Presentations* – connects to LibreOffice Impress or Microsoft PowerPoint and allows control of the presentation (navigating between slides, displaying the presentation notes directly in the application, etc.).

3. Power management options – options for locking the remote PC, rebooting it or turning it off.
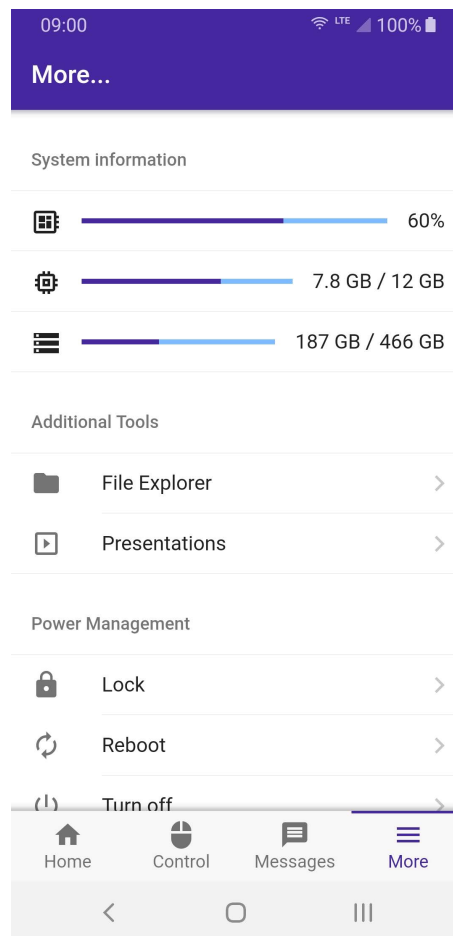
Figure 4: Screenshot of the additional options tab.

3   Technical Overview

3.1   Client Architecture

At its core, the client application was created using web technologies, making it consist of:

- HTML 5 (**H**ypertext **M**arkup **L**anguage)
- CSS 3 (**C**ascading **S**tyle **S**heets)
- JavaScript.

However, the development was not done in CSS or pure browser JavaScript, the HTML being the only one remaining as-is. The following technologies were used instead:

- *Less* (**L**eaner **S**tyle **S**heets) is a dynamic preprocessor stylesheet which transpiles to CSS 3 that can be interpreted by browsers but adds additional features. One of its most important features is nesting which allows for less verbose stylesheets, as well as variables and expression computation.

- *ECMAScript 2017* is a newer, standardized iteration of the JavaScript programming language, a version which is not yet fully supported by the latest browsers. As such, a tool called *Babel* takes the ES2017 source code and transpiles it into browser-compatible ES4. For a comparison of features for each of the ECMAScript specification, see Figure [10].

3.2   Server Architecture

The server component of the application is targeted to run on the PC that is being remotely-controlled. As a relay of information to the mobile application and command executor, it needs access to the native functions of the host operating system.

It was decided to use JavaScript to power the server, more specifically a more modern version of JavaScript called ECMAScript. However, JavaScript usually cannot run directly on a server as it was designed specifically for the web, to be run in a browser; to overcome this limitation, Node.js® was used.

Node.js® is a cross-platform runtime that uses Chrome's V8 JavaScript engine that can be used outside a browser and proves to be very suitable for running a JavaScript back-end[4]. It has an asynchronous event-driven architecture and can be used to build scalable network applications.

By itself, the Node runtime offers great native-like functionality which would not have been possible with the browser security level, such as: file system access, native operating system API calls, native plugins, etc. Considering the purposes of the remote control application, having access to the operating system and the file system was mandatory.

3.2.1   ECMAScript specification differences

Different versions of Node provide partial or full support for the different version of ECMAScript language specifications. For the purposes of this development, Node 11.15.0 was selected as it was the latest stable version at the time. This particular version of Node offers the following compatibility with the ECMAScript specifications:

Specification	11.15.0 Compatibility	Feature highlights
ES2015	99%	Modules, template strings, shorthand object literals, computed property names, fat arrow functions[2]
ES2016	Full compatibility	Exponentiation operator, array improvements, generator & destruturing improvements[3]
ES2017	Full compatibility	Asynchronous functions, shared memory and atomics[8]
ES2018	Full compatibility	Rest/spread properties, async iteration, promise improvements, regex improvements, template literal revision[6]
ES2019	87%	Additions in array & string processing, catch bindings, JSON superset[9]
ESNEXT	10%	Features proposed for the next ECMAScript release (i.e. *ES2020*), bound to change as specificatins mature

Figure 5: A short comparison of the different versions of the ECMAScript language specification.

Of these ECMAScript specifications, ES2015 and ES2017 proved instrumental in the development of the application as it allowed a modular approach in developing the server and an object-oriented approach that is more similar to the Java world than the prototype-based system ES4 has.

3.2.2   Client-server communication

In order to create the communication system between the client and the server, multiple possibilities were considered:

1. **SOAP** (*Simple Object Access Protocol*) is a data protocol that would allow the exchange of data between the client and the server, by using XML projects. This particular protocol was, at the time of its introduction, popular in the corporate world for its extensibility.

2. **REST** (*Representational State Transfer*) is an architectural style that usually runs over HTTP and has a strong presence in Web APIs due to its ease of use and easy integration with JSON.

3. **WebSocket** is a full-duplex communication protocol, allowing for real-time communication and is relatively new in the web world, being introduced in 2011.[5]

Since the REST architectural style is closer to the web world than the XML-based SOAP, it was selected as the primary communication method. The concepts of "web resources" and HTTP methods (`GET`, `HEAD`, `POST`, `PUT`, `PATCH`, `DELETE`, `CONNECT`, `OPTIONS` and `TRACE`) fit the status and command-based system a remote control application would have.

The remaining technology would be *WebSockets*, case in which real-time communication is not necessary for most uses of this particular application. However, in the end it was decided to use this technology as complementary for the REST architectural style in some specific cases which will be discussed further.

While HTTP servers are natively supported, there is a rather popular NPM (Node Package Manager) module that provides an HTTP server with native support for RESTful APIs called *Express*[3]. Taking a minimalist approach towards web server, Express allows definition of stand-alone routes which can be composed in order to create a fully-fledged API.

3.2.3 API description

The RESTful API follows the structure of the modules presented earlier, by adding resources which allows the client to send commands to or receive a status from the PC being controlled remotely pertaining to a specific module.

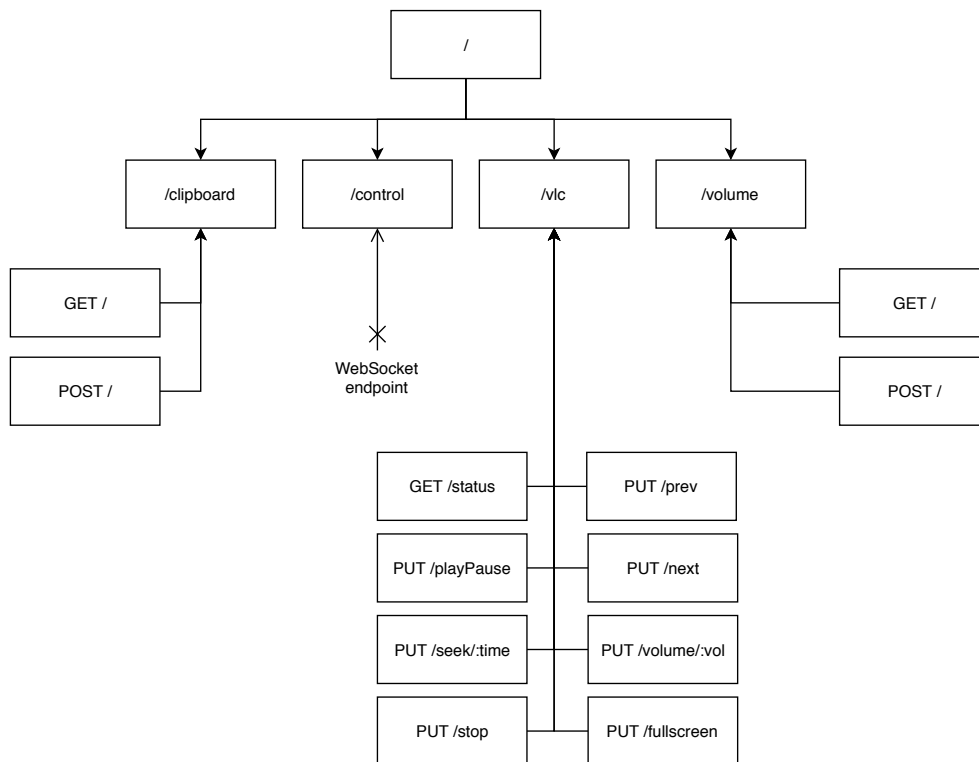Figure 6: A hierarhical diagram of all the RESTful routes exposed by the server. The nodes that have a leading slash are branches of the API; only the nodes that express the HTTP method (e.g. GET, POST) are accesible routes.

With the 6 presenting the overall structure of the RESTful APIs, an individual summary of each of the end points available to the user would be in order.

---

[3] https://www.npmjs.com/package/express

Method	URL	Description
GET	/clipboard	Returns a JSON response with the contents of the PC's clipboard.
POST	/clipboard	Sets the content of the PC's clipboard based on the value provided via the `application/x-www-form-urlencoded` format.
GET	/vlc/status	Returns the last known status of the VLC media player, in a simplified JSON format.
PUT	/playPause	Causes the VLC player to pause the current media if already playing, or resume it otherwise.
PUT	/prev	Causes the VLC player to go to the previous media in the playlist.
PUT	/next	Causes the VLC player to go to the next media in the playlist.
PUT	/seek/:time	Causes the VLC player to seek the current media to the position specified by the user (e.g. /seek/60 would seek to 00:01:00).
PUT	/volume/vol:	Causes the VLC player to change its volume to the specified value from 0 to 512 (e.g. /volume/256 would set the volume to 100%).
PUT	/stop	Causes the VLC player to stop the current playlist.
PUT	/fullscreen	Causes the VLC player to toggle its full screen status.
GET	/volume	Obtains the global system volume of the PC (usually 0-100).
SET	/volume	Sets the global system volume of the PC (usually 0-100).

Figure 7: The list of all endpoints exposed by the server's RESTful API, including their HTTP method and a short description

3.2.4 Real-time communication API

Apart from the standard RESTful API presented earlier, the server also has the capability of supporting modules which require bidirectional real-time communication between the client and the server.

As previously mentioned, for real-time communication requirements *WebSockets* would be used. In order to improve the reliability of communicating via TCP/IP and to make sure that the client and the server can communicate even if the client's browser does not support web sockets, a third-party solution can be used.

Socket.IO[4], according to its description "enables real-time, bidirectional and event-based communication. It works on every platform, browser or device, focusing equally on reliability and speed."[1]

Currently, only the *Control* module uses this particular real-time communication due to the high data throughput generated by the client sending mouse coordinates and keystrokes. In this case communication from the server back to the client is minimal since the remote control works unidirectionally.

The *Control* module works by sending from the client to the server two possible events/commands:

- `mouseMove` with two integer arguments representing the X and Y delta. Upon receiving this command, the server applies the delta against its known current mouse position, causing the pointer to shift by the given coordinates.

- `mouseClick` with the ID of the mouse button to be pressed. Upon receiving the command, the server simulates a mouse click at the current mouse position.

Because moving the mouse pointer on the client application's touchpad can generate a high quantity of data (more than 50 `mouseMove` commands in less than two seconds), the following actions are taken by the client:

- No `mouseMove` command is sent if the user has dragged across the touchpad for less than three pixels.

- There is a debouncing effect that causes multiple short drags to be sent as one, instead of generating multiple actions every time the finger is lifted off (this is especially important since touchpad detection can sometimes indicate multiple short interruptions in the drag gesture).

---

[4]https://socket.io/

3.3 VLC Remote Library

One of the main features of the application is the ability to control multimedia players and for the purpose of this implementation the VLC Media Player was chosen as the primarily supported application.

Unlike some graphical applications, this particular media player offers multiple graphical and non-graphical control interfaces, more specifically[5]:

1. Graphical interfaces:

    (a) **Qt interface**: graphical user interface that uses the cross-platform Qt toolkit

    (b) **Skinnable interface**: similar in functionality to *Qt interface* but offers the possibility of using custom themes (or skins) for the application.

2. Console/text-based interfaces:

    (a) **Ncurses interface**: a text-based user interface that can run in a console/terminal directly on the local machine or remotely over SSH.

    (b) **Telnet**: uses an unsecure Telnet connection to allow text transmission of the commands and status.

    (c) **Console**: the command line interface routed through the Lua scripting language

    (d) **Remote control interface**: another method of using the CLI or TCP/IP for remote control.

3. Non-standard interfaces:

    (a) **Web**: allows the remote control of the application via a web interface which can be accessed via the local network.

    (b) **Lua interpreter**: allows the use of scrips written in Lua (run via VLC's built-in interpreter) to control the application.

According to the official documentation[11], there are additional interfaces apart from the ones listed here. However, these particular interfaces are outside the scope of this article. At the time of writing, these interfaces not mentioned previously are:

Graphical: macOS, wxWidgets

Others: Hotkeys, lirc, osc and Wiimote.

Of these interfaces, the main contenders for a remote control interface would be *Web* or *Remote Control Interface*. Because the application was going to be developed in JavaScript using web technologies, it was decided to use the *Web* interface of the VLC Player. While using a web interface rather than a TCP/IP socket is arguibly less performant, the impact is minor considering the fact that the remote control usually has a low data throughput. Apart from using web technologies, a major factor in this decision was also the development time, as it was easier to use the existing web infrastructure than to communicate via sockets.

Despite the abundance of JavaScript modules from the Node Package Manager (NPM) repository, there was no suitable library for remote-controlling VLC, at the time of writing.

The following Node packages were taken into consideration when trying to find a dependency:

- `vlc`[6]: this particular library uses `libvlc` which would make the application act as a stand-alone multimedia player[12], rather than a remote-control application for VLC;

- `vlc-command`[7]: would only expose the path to the `vlc` executable;

- `vlc-remote`[8]: uses the TCP/RC protocol instead of the desired web protocol, `vlc-remote-alpha` seems to be the discontinued version of this library.

---

[5]The list of user interfaces was obtained directly from the application by going through Tools–Preferences–All (Show settings)–Interface–Main interfaces.

[6]`https://www.npmjs.com/package/vlc`

[7]`https://www.npmjs.com/package/vlc-command`

[8]`https://www.npmjs.com/package/vlc-remote`

Considering the apparent lack of a library, a custom Node package was needed. The first step in this direction was to research the workings of the *Web* interface.

The web interface must be enabled first from VLC's Preference window in order to work. Upon enabling it, restarting the application would create a web server on the local machine that can be used to access the built-in remote control application.

The default port is 8080 and considering the server is hosted locally, it can be accessed via the URL: `http://localhost:8080`. However, unless filtered out by the system's firewall, this particular port would also be exposed on the local network.

Upon accessing the webpage, an authentication screen is presented. The authentication screen is part of the browser, meaning that the web interface uses a form of HTTP authentication[7], most likely Basic Authentication. Since the configuration page of the interface only allows the user to add a password, the user name field should be left empty.

Using the incorrect credentials when authenticating will result in an error page returning HTTP status code 401 (Client error). A successful authentication would lead to the remote control page displayed in figure 9.
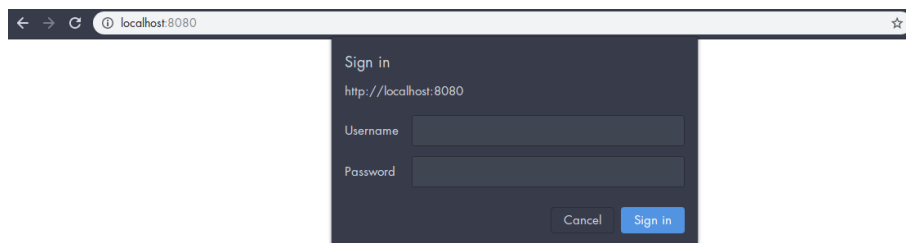

Figure 8: The authentication dialog presented to the user when first accessing the VLC web interface, on a *Google Chrome* browser.
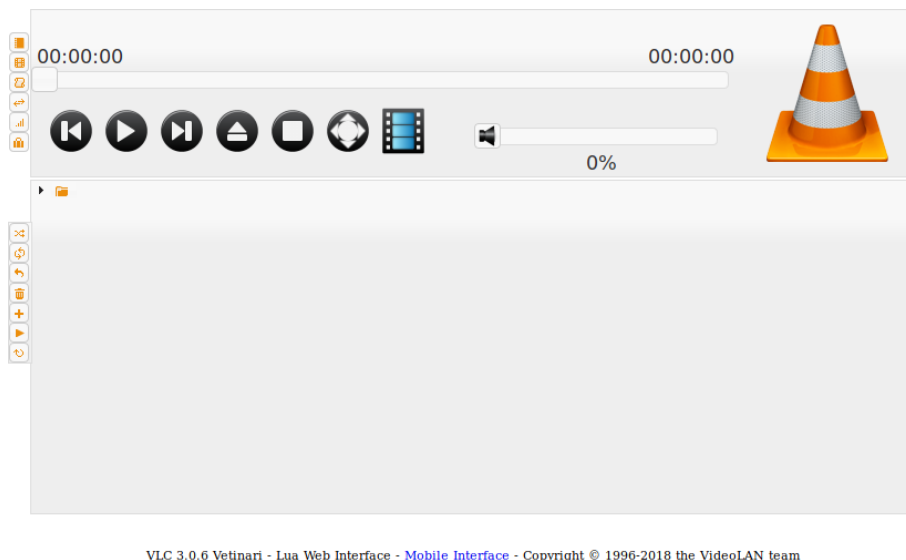

Figure 9: The default web interface of VLC, as seen from a *Mozilla Firefox* browser.

Inspecting the idle state of the remote control using Mozilla's Development Tools reveals that the way the web interface updates its status (the seek position, the file name, play state, etc.) is via polling. Polling[13], in this case, refers to making a periodical request to check for any changes in status.

The URL being polled is `http://localhost:8080/requests/status.xml` and this is done every 1000ms; the timing was most likely set to update the media time seek indicator.

As the file extension of the status request URL reveals, the format received is *eXtensible Markup Language*. The following listings are samples of the response received from the call for status:

Listing 1: "The request response to the /status route when there is a media playing in VLC"

```xml
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<root>

<audiofilters>
  <filter_0></filter_0></audiofilters>
<equalizer></equalizer>
<random>false</random>
<position>0.22149999439716</position>
<state>paused</state>
<audiodelay>0.0</audiodelay>
<time>6</time>
<videoeffects>
  <hue>0.0</hue>
  <contrast>1.0</contrast>
  <saturation>1.0</saturation>
  <gamma>1.0</gamma>
  <brightness>1.0</brightness></videoeffects>
<apiversion>3</apiversion>
<aspectratio>default</aspectratio>
<volume>0.0</volume>
<subtitledelay>0.0</subtitledelay>
<fullscreen>false</fullscreen>
<rate>1.0</rate>
<version>3.0.6 Vetinari</version>
<length>30</length>
<repeat>false</repeat>
<loop>false</loop>
<currentplid>4</currentplid><information>
    <category name="meta">
    <info name='filename'>Screencast from 02-12-2017 04:32:10 PM.webm</info>    </category>
  <category name='Stream_0'><info name='Orientation'>Top left</info><info name='Codec'>Google/On2&#39;s
      VP9 Video (VP90)</info><info name='Decoded_format'>Planar 4:2:0 YUV</info><info name='Frame_
      rate'>30.000300</info><info name='Buffer_dimensions'>3936x1088</info><info name='Video_
      resolution'>3926x1080</info><info name='Language'>English</info><info
      name='Type'>Video</info></category>   </information>
  <stats>
  <readbytes>1299231</readbytes>
<sentbytes>0</sentbytes>
<decodedaudio>0</decodedaudio>
<lostpictures>5</lostpictures>
<lostabuffers>0</lostabuffers>
<decodedvideo>302</decodedvideo>
<demuxdiscontinuity>6</demuxdiscontinuity>
<demuxreadpackets>0</demuxreadpackets>
<sentpackets>0</sentpackets>
<demuxbitrate>0.014810333028436</demuxbitrate>
<displayedpictures>302</displayedpictures>
<readpackets>96</readpackets>
<sendbitrate>0.0</sendbitrate>
<inputbitrate>0.0</inputbitrate>
<playedabuffers>0</playedabuffers>
<averageinputbitrate>0.0</averageinputbitrate>
<averagedemuxbitrate>0.0</averagedemuxbitrate>
<demuxreadbytes>6104132</demuxreadbytes>
<demuxcorrupted>0</demuxcorrupted>
  </stats>
</root>
```

Listing 2: "The request response to the /status route when there is no media playing in VLC"

```xml
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<root>
```

```
<loop>false</loop>
<equalizer></equalizer>
<random>false</random>
<position>0</position>
<state>stopped</state>
<audiodelay>0</audiodelay>
<time>0</time>
<videoeffects>
  <hue>0.0</hue>
  <contrast>1.0</contrast>
  <saturation>1.0</saturation>
  <gamma>1.0</gamma>
  <brightness>1.0</brightness></videoeffects>
<apiversion>3</apiversion>
<repeat>false</repeat>
<volume>0.0</volume>
<subtitledelay>0</subtitledelay>
<fullscreen>0</fullscreen>
<rate>1</rate>
<version>3.0.6 Vetinari</version>
<length>0</length>
<audiofilters>
  <filter_0></filter_0></audiofilters>
<currentplid>−1</currentplid><information>
    <category name="meta">
        </category>
    </information>
  <stats>
    </stats>
</root>
```

Apart from the status query, an important part would also be sending specific commands (previous/next media, play/pause, set volume, etc.).

Commands are sent by making another request to the status URL, with the difference of adding a query parameter to indicate the command to be sent. For example, to send a *Play/Pause* command a request to `http://localhost:8080/requests/status.xml?command=pl_pause` would be made. For commands that also require a value, such as as setting the volume a `val` query parameter is added.

The request response after sending a command is the same as a normal status query, meaning that if a call to a command was made there is no need to query the status again from the server, causing two different requests.

As required by the project, the following functions were implemented:

VLC command	Equivalent JavaScript function	Description
pl_pause	playPause()	Causes the player to pause the current media if already playing, or resume playing otherwise.
seek	seek()	Causes the player to seek to the specified time.
pl_previous	prev()	Causes the player to go back to the previous video in its playlist.
pl_next	next()	Causes the video to go to the next video in its playlist.
pl_stop	stop()	Causes the player to stop, usually closing the current stream entirely.
fullscreen	toggleFullscreen()	Causes the player to go into full screen, if not already activated. Otherwise, it will exit full screen. Going into full screen works regardless of the window focus or minimization state.
volume	setVolume()	Adjusts the volume of the player, with a number ranging between 0 and 512. A volume of 256 represents 100% volume, whereas 512 represents 200% amplified volume.

4 Conclusions

4.1 Future directions

This paper and its implementation proves that it is certainly viable to create an unified and extensible open-source remote control application. The experiment to use JavaScript both for the client application and the server also proved succesful since there were no noticeable performance issues with the implementation.

While the project has already gained traction in its development, in order to reach a higher level of competitivity and improve user experience new features would have to be added. The following points are a list of improvements that will be done gradually.

1. Adding more control widgets of general interest which are displayed on the home page.

2. Extend existing control widgets to add more specific functionality:

   - *General options widget*: add more audio-related options such as the ability to change the input/output devices and find additional options other than audio and brightness control.
   - *Media player widget*: add more control options such as previewing the playlist, displaying a thumbnail of the media being played, control subtitles and audio tracks.
   - *Control tab*: add live screenshot of the remote PC while using the touchpad, add support for functional keys for the virtual keyboard.
   - *Messages tab*: add support for more message preview types such as video files or documents.
   - *System information widget*: display battery state information (percentage, charging, estimated time left, etc.).
   - *Power management widget*: add support for more power actions such as Hibernate.

3. Improve compatibility with more devices or applications:

   - Application-wide support for macOS. Currently this platform is not supported.
   - *Media player widget*: add support for media player applications other than VLC Player.
   - *Presentation tool*: add support for other presentation applications than LibreOffice Impress or Microsoft PowerPoint.

Acknowledgment

References

[1] Automattic. *Socket.IO.* 2019. URL: `https://socket.io/`.

[2] Google Developers. *ES2015.* 2019. URL: `https://developers.google.com/web/shows/ttt/series-2/es2015`.

[3] Ralf S. Engelschall. *ECMAScript 6 — New Features: Overview & Comparison.* 2017. URL: `http://es6-features.org/`.

[4] Node.js Foundation. *About Node.js®.* 2018. URL: `https://nodejs.org/en/about/`.

[5] Internet Engineering Task Force (IETF). *Request for Comments: 6455.* 2011. URL: `https://tools.ietf.org/html/rfc6455`.

[6] Faraz Kelhini. *New ES2018 Features Every JavaScript Developer Should Know.* 2019. URL: `https://css-tricks.com/new-es2018-features-every-javascript-developer-should-know/`.

[7]   Mozilla Development Network. *HTTP authentication*. 2019. URL: `https://developer.mozilla.org/en-US/docs/Web/HTTP/Authentication`.

[8]   Dr. Axel Rauschmayer. *ECMAScript 2017 (ES8): the final feature set*. 2016. URL: `http://2ality.com/2016/02/ecmascript-2017.html`.

[9]   Selvaganesh. *JavaScript: What's new in ECMAScript 2019 (ES2019)/ES10?* 2019. URL: `https://medium.com/\@selvaganesh93/javascript-whats-new-in-ecmascript-2019-es2019-es10-35210c6e7f4b`.

[10]   Stoyan Stefanov. *JavaScript Patterns*. O'Reilly Media, Inc., 2010, p. 5. ISBN: 9781449396947.

[11]   VideoLAN Wiki. *Documentation:Modules*. 2019. URL: `https://wiki.videolan.org/Documentation:Modules/`.

[12]   VideoLAN Wiki. *libVLC*. 2019. URL: `https://wiki.videolan.org/LibVLC/`.

[13]   Wikipedia. *Polling (computer science)*. 2019. URL: `https://en.wikipedia.org/wiki/Polling_(computer_science)`.

Doran Adoris Elian

Lucian Blaga University of Sibiu

Faculty of Sciences, Dept. of Informatics

Str. Dr. I. Rațiu, No. 5-7, Sibiu 550012

România

E-mail: *elian.doran@ulbsibiu.ro*

ChessEngine

Răzvan Gheorghe Filea

Abstract

This work describes the implementation of a cross-platform algorithm for the chess game and its ongoing optimization after a careful analysis of the workflow and also due to the desire of continuously improving it. The game is developed as an Android application used to improve the chess knowledge of the players. It allows the player to play with the black and also with the white pieces. The application is simple to use, because it has a friendly user interface and it also permits settings regarding the used algorithm. It contains a statistical analysis regarding the used and improved algorithms during the development of the application.

1 Introduction

The application is an implementation of the classic chess game and it has been developed as a mobile application in Android Studio using the Kotlin programming language. The algorithm, which determines what moves are made, was written in C++ using the Visual Studio Community 2017 integrated development environment. The purpose of the application is to improve players' chess skills to make them chess players as valuable as possible. Other similar applications are: *Lichess*, *DroidFish*, *Komoda Chess Engine*. All the listed apps only implement the interface for an already created chess engine and unlike them, the presented application uses its own chess engine and allows deep customization of the algorithm's search parameters.

This work is structured into 7 chapters comprising the steps taken in the development of the application, namely *Introduction* that includes a brief appraisal of the application and its framing into the current context. Chapter 2 contains the *Application Description* as structure, interface and, functionality. Chapter 3 describes the data structure used for *Chessboard Representation*. Chapter 4 includes the *Algorithm Development* and the steps taken to develop and evolve it over time. Chapter 5 includes *The Evaluation Function* of possible moves. Chapter 6 contains details on the *Algorithm Optimizations* using algorithms like Alpha-Beta Pruning, Zobrist Hashing and Late Move Reductions, all of these used to obtain better performance. Chapter 7 contains the conclusions of the work and proposals for further developments and improvements.

2   Application Description

The application opens with the main screen containing the chess board, a game restart button and a button for accessing the settings: Fig. 1. On the main screen, the user can start playing with white pieces while the black pieces are normally handled by the application. Although, when restarting the game, the player may choose which side he wants to play on. The pieces are represented by images downloaded from the Wikimedia Commons website [6].

Fig. 1: Main screen of the application

A piece may be selected by pressing on it. To move it, one may press one of the squares that have turned green. These represent the valid movements that can be performed by that piece. The last moved piece is marked by yellowing the start and destination squares of that piece. To signal the important events in the game, namely when a player is in check or checkmate, that player's king is marked in red. And if the game is over, a dialog will be displayed to inform the user of this.

After a move has been made by the user, the application performs an external function call, i.e. a special C++ written function that can be called by the JVM (Java Virtual Machine). This function creates a new thread so it does not block the app's Main Thread and starts the search. This new thread waits for the search to complete while it is performed in parallel by multiple threads and then notifies the app. The search is completed when all possible moves have been examined at the user-specified depth in the settings.

When the application is closed all the moves that have been made are saved into a file. The file is named *"moves.sav"* and is located in the internal storage, meaning it can only be accessed by the application that created it. When the application is started again, it uses the stored moves from the file to restore the last game state, allowing the user to continue their game at another time.

The colors of the squares on the board, the color of the possible movements, the color of the selected piece, the last movement, and the color of the king in check can all be customized from the settings, Fig. 2.

The "Base Search Depth" setting determines the difficulty of the algorithm, as it represents the depth of search for possible moves. Since the computer does not have a strategy, unlike a person who develops one over time, the depth of the search plays a very important role in how good the "strategy" of the algorithm is. On identical execution scenarios, with no optimizations, assuming that each move performed shows an average of other 35 possible moves, the algorithm must look for about 35^n moves, where n represents the specified search depth.

The "Thread Count" setting determines the number of threads used by the search algorithm. This is implicitly set to the total number of CPU cores - 1, and the maximum number at which it can be set is the total number of CPU cores. Setting the number of threads to the highest number will result in finding the best possible move faster. For example, on similar execution scenarios and with a search depth of 4 moves in both cases, using a single thread the algorithm needs approximately 520 milliseconds to move, whereas using 4 threads it only needs 300 milliseconds.
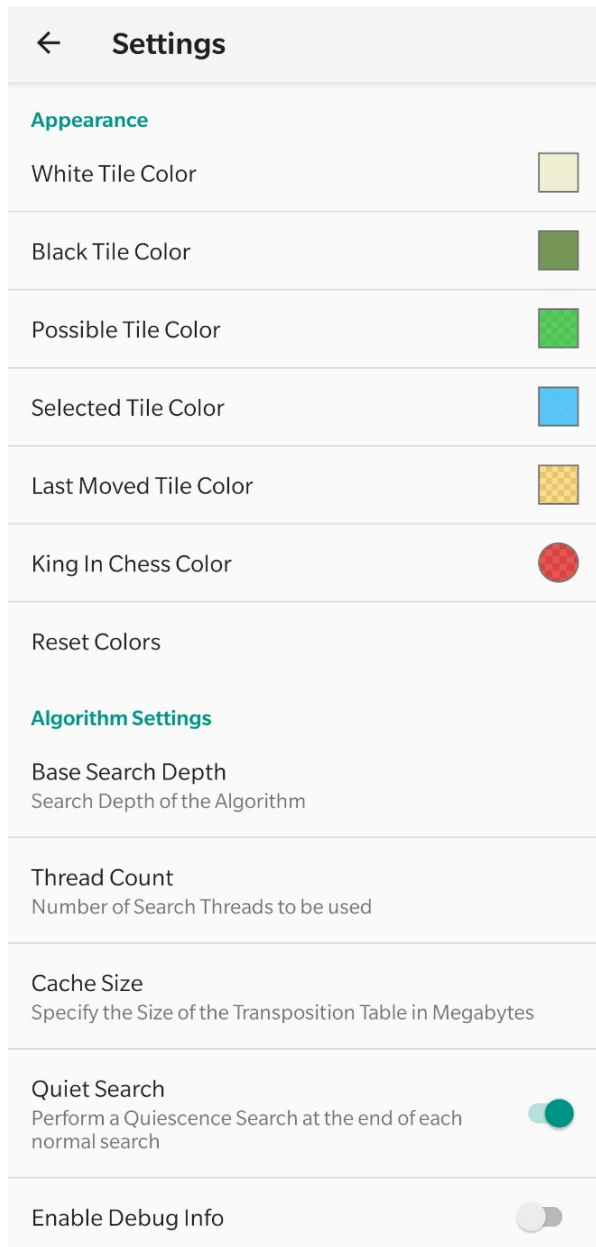
Fig. 2: The settings screen

The "Quiet Search" switch enables the use of the Quiescence Search Algorithm, which can, especially with small search depths, considerably increase the overall difficulty of the algorithm. This algorithm will later be described more thoroughly.

The "Cache Size" setting determines how much memory is used by the transposition table. Increasing this value might increase the search speed since it can store multiple search nodes. Although, the default value of 200MB should be enough for short searches.

Turning on the "Enable Debug Info" switch will display algorithm statistics below the chessboard after each move, such as: the number of moves evaluated, the number of nodes searched, the exact time required to perform the search, the number of dynamic allocations made, and the rating of the current board from the perspective of the algorithm.

3 Chessboard Representation

The representation of the chess table is made through a data structure called *Board*. The most important part of this data structure is the array called *data* because it contains the pieces on the chessboard. At first, this variable was defined as an array of 64 pointers to the pieces:

*Piece *data[64];*

Piece was an abstract class, each actual piece being a derived class and one's absence from the array was represented through the value *null*.

Once one has realized that this way of representing the chessboard is very inefficient, one has decided to store the pieces in continuous memory even if it meant giving up the polymorphism. Also, in order to make it easier to access a specific position on the board, a matrix of 8x8 pieces has been decided to be used:

Piece data[8][8];

Other variables used in the Board data structure can be seen below, along with a description of their role:

```
unsigned long long key; // The key generated by the hashing function.
// It is needed for the transposition table.

bool whiteCastled, blackCastled; // If the white, respectively black
// player has castled. It is used by the evaluation function.

State state; // Indicates if one of the players is in check or checkmate.
bool whiteToMove; // Indicates if it is the white player's turn
int score; //The score of the board, calculated by the evaluation function
```

4 Algorithm Development

The algorithm is written in C++17 and from the very beginning, the application was developed to separate the interface and the algorithm. The fact that they are made completely separate, has ensured that the algorithm is cross-platform.

The development of the application started by sketching the basic data structures and later by deciding the structure of the application code. Since the initial sketching, all of the data structures have been rewritten at least once. But looking back it was still a good idea as it helped to keep the code organized.

After the move generation for each piece was done, a simple evaluation function was written. For the start, the evaluation function only added the score of the white pieces and subtracted the score of the black ones. Then it went on to develop the algorithm itself, which chose the best move. After long documentation, the Minimax [3] algorithm was decided to be implemented.

Minimax is a backtracking algorithm used to find the optimal move for a player, assuming that his/her opponent is also playing optimally (Fig. 3). In Minimax there are two players one trying to get the highest possible score, the player with white pieces in our case, and the other trying to get the lowest possible score, i.e. the player with black pieces.
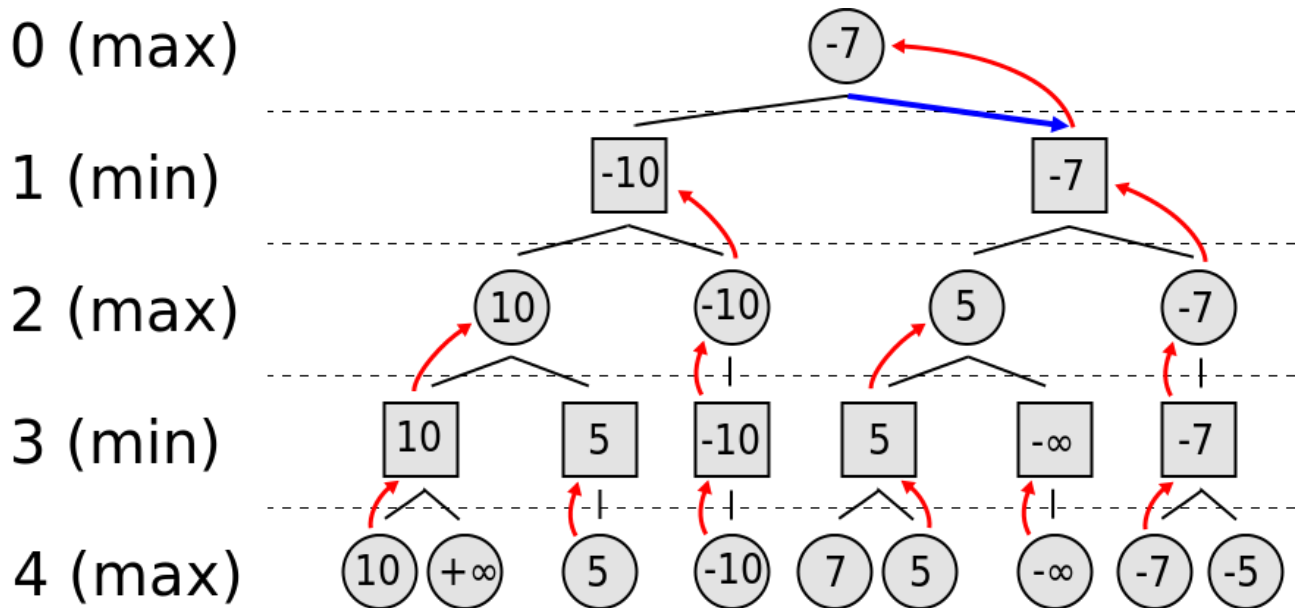
Fig. 3: A Minimax tree example [3]

Later, to simplify the implementation of the search algorithm, it was decided to implement the Negamax Search algorithm [3], a different form of the Minimax, that relies on the fact that *max(a, b) = -min(-a, -b)*. A simple implementation of this algorithm can be seen below:

```
int negaMax(const Board &board, short depth, bool isWhite)
{
    if (depth == 0)
        return isWhite ? board.score : -board.score;

    const auto validMoves = board.listValidMoves(isWhite);
    int bestScore = -VALUE_INFINITE;

    for (const Board &move : validMoves)
    {
        const int moveScore = -negaMax(move, depth - 1, !isWhite);
        if (moveScore > bestScore)
            bestScore = moveScore;
    }
    return bestScore;
}
```

This function first checks if the maximum depth is reached and if so, for Negamax to work properly, the score is reversed if necessary. Then it calls a function that returns a list of all valid moves and iterates through each element. For each move in the list, the function calls itself with the inverted color and the lowered depth. If the score of one of the moves is greater than the best move, the new score is chosen as the best move. Finally, the function returns the best score.

In order to simplify in the beginning the search algorithm, at first its purpose, i.e. getting the highest possible score, was limited to capturing the opponent's king. With the completion of implementing the algorithm bases and finishing some optimizations for it, it has been decided to implement an algorithm that complies with the basic rules of chess. This means that after each move, the algorithm has to check if one of the players is in check or checkmate. In order to check if one of the players is in check, the algorithm generates a list of all of the opponent's attacks and it checks if the king's position is in that list. And to check if it is checkmate, the algorithm checks whether the player has at least one valid move available, i.e. a move where it is not in check. Also, if one player cannot make any valid move but is not in check and it is it's turn to move, it is considered a stalemate, or more commonly known as a draw.

5   The Evaluation Function

The search algorithm is a key part of any chess program because it determines what is the best move in a certain context, but the evaluation function is equally important. It determines how good a position is and summarizes everything in a single integer variable so that it can be easily processed by the search algorithm.

At first, the evaluation function only considered the number of pieces on the board and their respective position. But it has become more and more complex regarding more and more aspects of the game board, such as piece mobility, doubled pawns, isolated pawns, passed pawns, rook on open file, high-mobility moves, pieces attacked by protected pawns, queen attacked by weaker pieces, bishops on long diagonals and others. All these criteria being evaluated differently depending on the game period.

Other "bonuses" are also granted to encourage certain practices. For example, the player will be rewarded if it is his turn to move, if he has castled, if he still has the chance to castle or is not in check. Because finding very good parameters for evaluating some aspects of the chessboard requires a constant adjustment of these values over an extended period of time, some values of the Stockfish Chess Engine [2] have been used.

6   Algorithm Optimizations

From the very beginning, an efficient, yet quick, algorithm implementation was attempted. In order to optimize and improve the algorithm, many other "helper" algorithms and also specialized data structures have been implemented.

An absolutely necessary algorithm is *Alpha-Beta Pruning* [3], a technique that reduces the number of nodes examined by the Minimax algorithm. It will stop the search once the variable *alpha*, the best possible move by white, is greater than or equal to the variable *beta*, the best possible move scored by black.

One of these specialized data structures is the *Transposition Table* [1]. Usually implemented as a Hash Table, it is used to speed up the game tree search by saving the result of a part of the already searched tree. Often the computer has to analyze a position several times. To avoid this problem, after the program meets a new position and analyses it, it stores it in the transposition table. Thus, the table is a sort of cache for the positions already analyzed.

Zobrist Hashing [1] is a hashing function that is essential for using a Transposition Table since it determines the hash of a position so it can easily be stored in memory. The reason Zobrist

Hashing is used instead of any other hashing function is that after the initial position hash has been computed, all other positions can be updated incrementally without having to recalculate the entire hash, which of course helps performance-wise.

Below it is demonstrated how a hash is incrementally updated after a move has been made:

```
// We remove the selected piece from the selected board position
key ^= Hash::getHash(selectedPos, selectedPiece);

// We remove the destination piece from the destination position, if it exists
if (destPiece) key ^= Hash::getHash(destPos, destPiece);
// Then we add back the selected piece on the destination position
key ^= Hash::getHash(destPos, selectedPiece);
```

Another program optimization that helped improve its performance is using multiple threads. My implementation of Negamax uses them to look for different branches of the tree in parallel.

An enhancement to the Alpha-Beta algorithm implemented in the chess engine is called *Late Move Reductions* (LMR) [4] and it attempts to examine a game search tree more efficiently. It uses the assumption that good move ordering causes a program to search the most likely moves early. The implementation used searches the first 4 moves at the full specified depth and then the rest of the moves that meet the conditions are searched at a reduced depth. Some of the conditions are that no player is in chess, the move is not a capture, the move is not a pawn promotion, the remained depth is greater or equal to 3 and that the current move was not already reduced.

Quiescence Search [5] is a more limited search, usually only examining captures, which is performed at the end of a normal search and is needed to avoid the horizon effect. Simply stopping the search when the desired depth is reached, is very dangerous. Consider the situation where the last move considered is the queen capturing the pawn. If the search stops there, the algorithm might think that it won a pawn. But what if it were to search one move deeper and find that the next move the queen is captured by another pawn. In this case, the algorithm actually lost a queen instead of winning a pawn. Alongside *Alpha-Beta Pruning* [3] and *Delta Pruning* [1], two techniques that greatly reduce the number of moves searched by Quiescence Search, this search algorithm is used to increase the difficulty of the chess engine.

An important improvement of the search algorithm are extensions namely, check extensions. If one of the players is in chess and the search reaches the specified maximum depth, the search depth is extended by 1. In order to not continuously extend the search, it can only be extended once. Check extensions are very easy to implement but can greatly help in finding the best move by the search algorithm.

All these improvements from the data structures to the algorithms used, as well as the use of multiple execution threads, have resulted in progressive optimization of the program.

7 Conclusions and Future Developments

By developing the application on a long-term basis, I have gradually learned to create thread-safe complex data structures and how to improve the implemented algorithms by understanding their operation in great detail. I also learned how the Java/Kotlin code, from an Android application, interacts with C++ written code.

The implemented algorithms can be further improved by using more algorithms or techniques that help reduce the number of positions examined. Possible analyzed algorithms are NegaScout, Killer Moves, and History Heuristic, many of these algorithms are explained on the Chess Programming Wiki [1].

To further increase the functionality of the application, I plan on saving each played game in a separate file containing all the moves of that game alongside with who won the game. Having a match history can help the player improve himself by reviewing what and when he made a mistake.

References

[1] Chess Programming Wiki, explains multiple used algorithms as well as the transposition table, 8 April 2019 - https://www.chessprogramming.org/

[2] The Stockfish GitHub Page, the use of some values for the evaluation function, 15 March 2019 - https://github.com/official-stockfish/Stockfish

[3] Wikipedia, explains the Minimax, Negamax and Alpha-Beta Pruning algorithms, 20 February 2019 - https://en.wikipedia.org/wiki/Minimax

[4] Mediocre Chess: LMR, explains the Late Move Reductions algorithm, 20 April 2019 - http://mediocrechess.blogspot.com/2007/03/other-late-move-reduction-lmr.html

[5] Chess Programming Wiki: Quiescence, explains the Quiescence Search algorithm, 16 April 2019 - https://www.chessprogramming.org/Quiescence_Search

[6] Wikimedia Commons, the images used to represent the chess pieces, 10 November 2018 - https://commons.wikimedia.org/wiki/Category:SVG_chess_pieces

[7] Dr. Kris Jamsa and Lars Klander "Totul despre C și C++", book used to learn the basics of the C++ programming" """"language

Răzvan Gheorghe FILEA
"Samuel von Brukenthal" National College
Sibiu, ROMÂNIA
E-mail: razvan.filea@gmail.com

Detecting Structural Variants in Tomato Genomes

Erika Goetz, Angelo Williams, Cooper Bodary, Sofia Visa, Esther van der Knaap

Abstract

This work illustrates the computational steps for discovering structural variants (such as deletions, inversions, duplications) in tomato genomes. Our approach is validated by the phylogenetic tree built from the structural variants. Further, the phylogenetic tree reveals the evolution of various tomato species. The main contribution of the authors is the design and implementation of filters and the use of a new and large genomic data set provided by our collaborators. This bioinformatics work is interdisciplinary and illustrates the applied side of computer science in genomics.

1   Introduction

Though they originated in the South American continent [7], tomatoes have become an important food source all over the globe. But not all tomatoes are created equally. Farmers are interested in selecting varieties with desirable properties such as a large size, a firm structure, and a juicy core. Properties such as these are regulated by structural variants (SVs) including insertions, deletions, inversions, and duplications of DNA segments in or around genes. Thus, identifying relevant SVs is important as it helps breeders create larger and tastier tomatoes. However, finding these SVs is a challenging computational task as illustrated below. In this paper we outline a two-step computational approach for discovering SVs using Illumina short paired-end reads of 245 tomato varieties.

The remaining part of the paper is organized as follows: the second section explains the biological data set and some basic Illumina sequencing, the third section presents the computational approach used to identify SVs, and the fourth section illustrates the use of SVs in biological analyses, such as building an SV-based phylogenetic tree to infer kinship and domestication hypotheses.

2   Basics of Illumina sequencing and the DNA data set

This section gives a brief overview of the Illumina sequencing technology that is used to generate our data set of 245 tomato varieties and presents the distribution of these varieties into ten classes according to their domestication level and geographical provenance.

2.1 Obtaining the DNA data with Illumina technology

This section describes the process of obtaining the DNA data using the Illumina DNA sequencing [6], the current technology for DNA reading. Briefly, the Illumina machine segments a large DNA string into many small fragments. Because the machine can only read the two ends of a DNA fragment, the middle of the fragment remains unknown resulting in the name short paired-end reads (or paired-ends). In our case, the fragments are of length 450bp (base-pairs), and Illumina reads 150bp from each end, leaving the middle 150bp unknown. Fig. 1 illustrates several such short paired-end reads with ends of 35bp and the middle unknown sequence of 330-430bp.
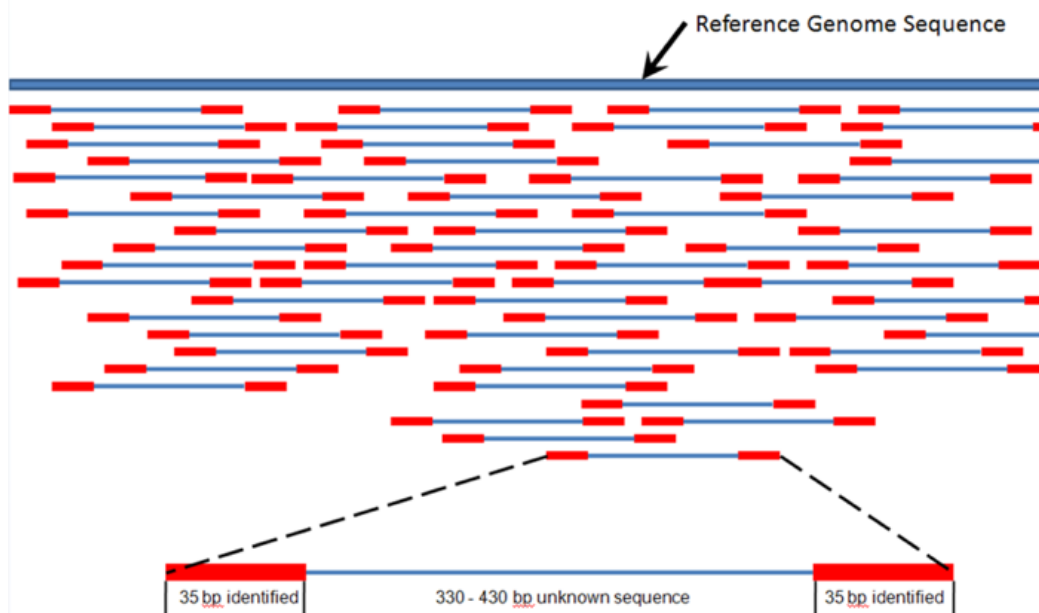

Fig. 1: Illustration of Illumina short paired-end reads aligned to a reference genome. [S. Martin, https://www.biostars.org/p/104218/]

To identify the position of the paired-end reads within one of the 12 tomato chromosomes, the two ends of a fragment are mapped to a reference genome which is already completely sequenced. Since 150bp is a short DNA string, each of the two ends might map in several locations within a reference genome, but we can choose the correct location based on the proximity of the two ends (i.e. the two ends should map at a distance of about 150bp of each other). Fig. 1 illustrates many Illumina paired-end reads aligned to a reference genome.

Since this process can lead to uneven coverage of the genome, it is a good idea to re-sequence the same genome several times to ensure that the entire genome is covered by at least one such small paired-end read. Fig. 2 illustrates a 5x coverage of a genome, meaning that the genome was sequenced with the Illumina technology several times, and on average we expect each nucleotide to be covered by five paired-end reads. Of course, due to the uncertainty of the method, this is not true for each nucleotide. For instance, a nucleotide from the middle of the

reference genome in Fig. 2 is covered by only one paired-end read. This information will be used later when we develop our computational pipeline and create filters for SVs.
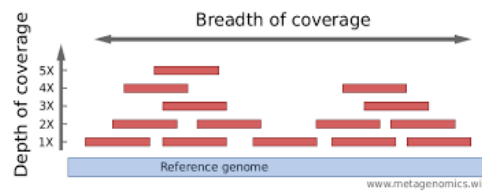

Fig. 2: Illustration of depth of coverage (or read depth). [www.metagenomics.wiki]

2.2 The DNA data set for our 245 tomato accessions

Using the technology outlined above, the DNA short paired-end reads for 245 accessions (i.e. tomato types) are obtained. Using the bwa software [4] these reads are aligned to the Heinz reference genome [3]. The depth of coverage varies across the samples from 8x to 45x, so our computational pipeline must account for coverage variation at the filtering stage.

Table 1 shows the distribution of the 245 tomato accessions into ten groups according to the provenance region. In addition, it shows the domestication stage of each group (see first column of the table); namely SLL (*Solanum lycopersicum* var *lycopersicum*) corresponds to the modern, domesticated tomatoes that we find in markets, SLC (*Solanum lycopersicum* var *cerasiforme*) corresponds to the semi-domesticated tomato, and SP (*Solanum pimpinellifolium*) represents the wild tomato. Representatives of these three domestication stages are illustrated in Table 2.

Geographically, the tomato varieties used in this research come from Mexico and the rest of the world (Major); Mexico (MEX); south Mexico, Central America, and Northern South America (MEX-CA-NSA); San Martin island (SAN MARTIN); Peru (PER); Ecuador (ECU); north Ecuador (NECU); south Ecuador (SECU).

Table 1. The distribution of the 245 tomato accessions into ten groups.

Domestication stage	Geographic group	Number of tomato varieties
SLL	1) Major	57
	2) MEX	14
SLC	3) MEX	29
	4) MEX-CA-NSA	16
	5) SAN MARTIN	13
	6) PER	27
	7) ECU	46
SP	8) NECU	10
	9) SECU	4
	10) PER	29

Table 2. Illustrations of SLL, SLC, and SP tomatoes.

SLL tomatoes	SLC tomatoes	SP tomatoes	SLL Heinz tomatoes

3   Detecting SVs in 245 tomato varieties

This section exemplifies three types of SVs and describes the main two steps for discovering SVs: SV calling and SV filtering. To call structural variants, we use an existing bioinformatics tool called LUMPY [1], and to filter structural variants, we wrote and used a series of Python scripts.

3.1 Illustration of structural variants

Genetic structural variants include insertions, deletions, inversions, and duplications. Fig. 3 illustrates three of these SVs. For instance, a deletion in a tomato genome of interest relative to the reference (i.e. another tomato genome, in our case the Heinz genome) is illustrated in the middle of the figure. This simply means that the newly sequenced genome is missing a chunk of DNA, but this can also be viewed as an insertion in the reference.

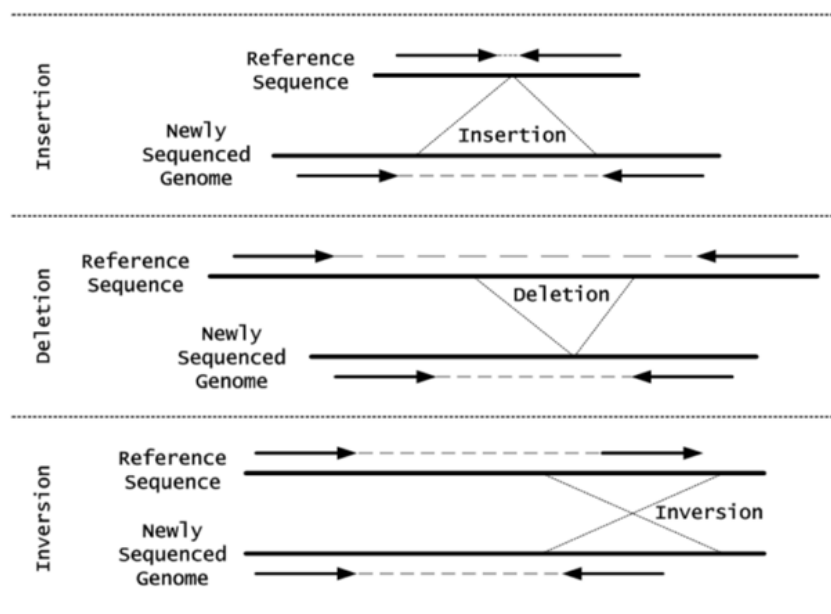Fig. 3: Illustrations of insertion, deletion, inversion SVs. [T. Rausch, EMBL]

3.2 Using LUMPY to call SVs

LUMPY is a probabilistic approach that uses short paired-end reads, split-reads, and read depth to call SVs. However, as explained in [1], the SV calls produced by LUMPY have many false positives and must be filtered according to additional information unique to each research project.

In this later stage lays our main contribution: we wrote Python scripts incorporating our own thresholds to filter these SVs as explained in the next section.

With the LUMPY software we call deletions (DEL), inversions (INV), duplications (DUP), and break ends (BND). BNDs calls need further analysis as they are a catch-all label for DNA fragments for which the two ends of a paired-end read map to two different chromosomes (note, the tomato genome has 12 chromosomes); for instance, such an event can indicate translocation.

LUMPY is called individually for each of the 245 bam files containing the aligned short paired-end reads. (A bam file is a standard bioinformatics binary file for storing sequences data [9].) This entire data set is large (about 2 terabytes) and computationally intensive. The SVs are called relative to the Heinz reference tomato genome, and this step outputs 245 VCF files (Variant Call Format, a standard bioinformatics file format [5]) which can be viewed with Excel. Each row in these files corresponds to an SV and lists information about it such as: its type, position, length, supporting evidence, etc. The supporting evidence (SE) is further used in the filtering stage of this analysis and consists of number of paired-ends (PE) and split reads (SR). By default, LUMPY calls an SV with as little as 4 total SE, but for a genome of 40x depth coverage, this constitutes little evidence as on average we expect 40 SE or even 20 SE, but not as little as 4 SE. Thus, in the filtering stage we will be using SE and depth coverage to filter out some false positives.

3.3 Filtering SVs

Since we use the default parameters of LUMPY (i.e. a minimum of 4 SE, regardless of read depth), we expect many false positives to be present among the structural variant calls that it made. To filter out such calls we implement Python scripts that read each of the 245 VCF files, filter them according to several criteria, and output 245 filtered VCF files, ready to be used in further analysis. The filtering criteria are as follows.

1) **Remove chromosome 0.** When there is ambiguity about the mapping location of a paired-end read, LUMPY lists it as belonging to an imaginary chromosome zero. Thus, first we remove all the entries in the VCF files listed as being located in this chromosome.

2) **Filter by evidence.** As briefly pointed out above, here we use the SE (=PE+SR) information relative to the read depth (d) of each of the 245 accessions to eliminate more false positives. Equations (1) and (2) show that we only keep SV calls that have at least one piece of each paired-end (PE) and at least one piece of split-read (SR) evidence. Further, equation (3) filters out SVs having more SE than three times the read depth (these can be transposons, which are not useful in this analysis); it also filters out SVs that have an amount of SE evidence less than half of the read depth. These thresholds are computed based on *apriori* biological information and on a small set of seven SVs computationally identified in the same way, which were later validated in lab by our collaborators.

$$PE >= 1 \tag{1}$$

$$SR >= 1 \tag{2}$$

$$3*d >= PE+SR > max\{4, d/2\} \tag{3}$$

3) **Filtering by "N".** Since the SL2.5 Heinz reference genome is not perfect and in some regions has long stretches of "N" (meaning that the correct nucleotide is unknown), we filter out SVs with start and/or end points in poorly mapped genome defined as the number of "N" being larger than 10 within 50 bp of start/end of the SV.

Table 3 illustrates the three stages of filtering for three of the 245 tomato accessions. As the table shows, after filtering there is a large reduction in SV counts. Following intuition, this

analysis expects a wild tomato (SP on last row) to be more distant from the Heinz tomato (which is an SLL) than the cultivated tomato (SLL on second row); hence an SP variety has more SVs (4,934) than an SLL one (714).

Table 3: Illustration of filtering SVs for 3 of the 245 tomato genomes.

Tomato name	No. of unfiltered SVs	No. of SV's after removing chr. 0	No. of SV's after removing by evidence	No. of SV's after removing by "N" criteria
SLL 7860	28,424	20,040	1,272	714
SLC 5912m	25,640	18,500	2,000	1,552
SP 6208	34,393	29,057	5,863	4,934

4 SV-based phylogeny tree of the 245 tomato varieties

Here we point out several further analyses one can perform with SV calls found above. First, these SVs can be overlaid with known existing genes of known phenotypes to eventually learn more about functionality. Second, various SV clustering across the 245 accessions can be done to discover common and unique genes. Third, one can build a phylogenetic tree to discover kinship and evolution (or domestication) trends, which we illustrate below. Notably, the SV-based phylogenetic tree obtained here confirms our SV detection method by grouping the 245 varieties according to their ten original groups.

To build the phylogenetic tree we wrote Python and MATLAB scripts that followed the steps below. For this analysis we used only DEL/INV/DUP (we did not use BND).

1) From the 245 filtered VCF files we created one masterfile having the union of all SVs in the 245 files. The master VCF file contains 41,469 unique SVs across all accessions. Each row in this file corresponds to an SV and, in addition to other information, it lists all the accession names having this particular SV. Note that in this masterfile we consider two SVs from two distinct tomatoes x and y as being the same SV if the two SVs

have the same type and
are in the same chromosome and
start at exact same position and
have same length.

Below, we relax the start position and length criteria to allow for the uncertainty of the SV calls which is a probabilistic approach. For instance, if in tomato x a DEL SV of length 1,000 starts at position 500 in chr. 2, then most likely it is the same as the DEL SV in tomato y starting in chr. 2 at position 502 and of length 1,001 or of length 996. This is an important observation that noticeably reduces the number of unique SVs as shown below.

2) Further, we filtered out SVs based on length and start/end proximity of an SV in two or more accessions.

We remove SVs larger than 1mil bp in length (see Table 4) and SVs smaller than 20bp.

We define an SV as being the *same SV* for two distinct tomatoes x and y if the two SVs
have the same type and
are in the same chromosome and
start and end at same location +/- 5 bp on each side.

This new definition of "same SV" creates another master file of 34,980 unique SVs (a major reduction from the 41,469 above).

Table 4: The length distribution of the 41,469 unique SVs; the SVs larger than 1 million bp are more likely false positives and are removed.

Length of SV	No. of SVs
1-100bp	2,557
100-1k	22,103
1k – 10k	10,958
10k-100k	3,919
100k – 1m	1,015
1m – 10m	492
10m – 50m	354
>50m	71

3) Next, we create a phylogenetic tree from a matrix with 245 rows (tomato varieties) and 34,980 columns (SVs). This is a 0/1 matrix, with an entry of 1 indicating that a particular tomato variety (row) has a particular SV (column).

To build the phylogenetic tree in MATLAB [8] we use the above matrix and cluster using Jaccard distance and averaged cluster linkage. Jaccard distance, which measures dissimilarity, is defined in equation (4), and it is basically one minus the intersection over the union of the SV sets of two tomatoes x and y. For instance, if the two tomato varieties have the same 100 SVs, the distance between them is 1-100/100 = 0, and thus they will be clustered together as they are very close (in fact, identical from the SV perspective).

$$Distance(x, y) = 1 – (SVs \ of \ x \cap SVs \ of \ y) / (SVs \ of \ x \cup SVs \ of \ y) \qquad (4)$$

The resulted phylogenetic tree of the 245 accessions is very large and it is shown in the appendix, together with four zoom-ins of it so that one can analyse it closer. An aggregation of this larger tree is shown in Fig. 4. This tree shows that the 10 classes from Table 1 segregate nicely, validating that the SVs we discovered are trustworthy and thus can be used in further analyses. This tree can also be used to infer some hypotheses regarding kinship and domestication. For instance, one can hypothesis that SLC ECU was domesticated from ancestors in the SP NECU group. Further the tree reflects the fact that the SLC SAN MARTIN group evolved in isolation which makes sense since these varieties grow on an island.
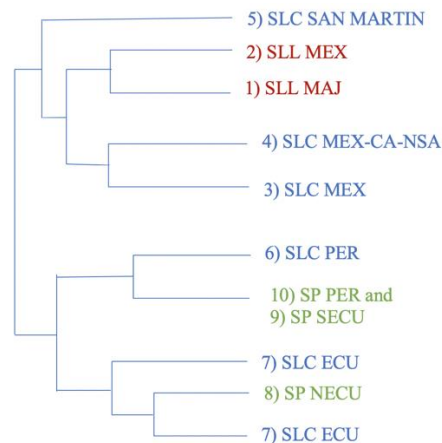
Fig. 4: Aggregation of the phylogenetic tree from Fig. 5. In red, blue, and green are the domesticated, semidomesticated, and wild tomatoes, respectively.

5 Conclusions

Here we illustrate the use of the bioinformatics tool LUMPY for structural variants detection, and we show that computational filtering leads to a valid subset of SVs. In addition, we illustrate one of many ways in which these SVs can be used in genomic analyses. Namely, we built an SV-based phylogenetic tree which shows the evolution of the 245 tomato species. This tree reveals that the first domestication happened in North Ecuador and that the modern tomato originates from Mexico. In future work we will be using these SVs to identify new size and taste genes in tomato fruits.

References

[1] R. M. Layer, C. Chiang, A. R. Quinlan, and I. M. Hall. *"LUMPY: A Probabilistic Framework for Structural Variant Discovery."* Genome Biology 15.6, 2014.

[2] A. R. Quinlan, and I. M. Hall. *"BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features."* Bioinformatics 26.6: 841-42, 2010.

[3] The Heinz tomato genome SL2.5, https://solgenomics.net/organism/Solanum_lycopersicum/genome, [Accessed in 2018].

[4] H. Li and R. Durbin. *Fast and accurate short read alignment with Burrows-Wheeler Transform.* Bioinformatics, 25:1754-60. [PMID: 19451168], 2009; bwa, Burrows-Wheeler Aligner tool, http://bio-bwa.sourceforge.net. [Accessed in 2018]

[5] Variant call format, vcf file format. http://www.internationalgenome.org/ [Accessed in 2019]

[6] Illumina DNA sequencing technology, https://www.illumina.com/techniques/sequencing [Accessed in 2019]

[7] J. Blanca, J. Montero-Pau, C. Sauvage, G. Bauchet, E. Illa, M.J. Diez, D. Francis, M. Causse, E. van der Knaap, J. Canizares. *Genomic variation in tomato, from wild ancestors to contemporary breeding accessions.* BMC Genomics 16, 257, 2015

[8] MATLAB Phylogenetic tree. https://www.mathworks.com/help/bioinfo/phylogenetic-analysis.html [Accessed in 2018]

[9] Bam file format, https://genome.sph.umich.edu/wiki/BAM [Accessed in 2019]

Erika GOETZ
College of Wooster
Computer Sci. Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: egoetz20@wooster.edu

Angelo WILLIAMS
College of Wooster
Computer Science Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: awilliams21@wooster.edu

Cooper BODARY
College of Wooster
Computer Science Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: cbodary19@wooster.edu

Sofia VISA
College of Wooster
Computer Sci. Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: svisa@wooster.edu

Esther van der KNAAP
Univ. of Georgia
The Plant Center
111 Riverbend Rd
Athens GA 30602
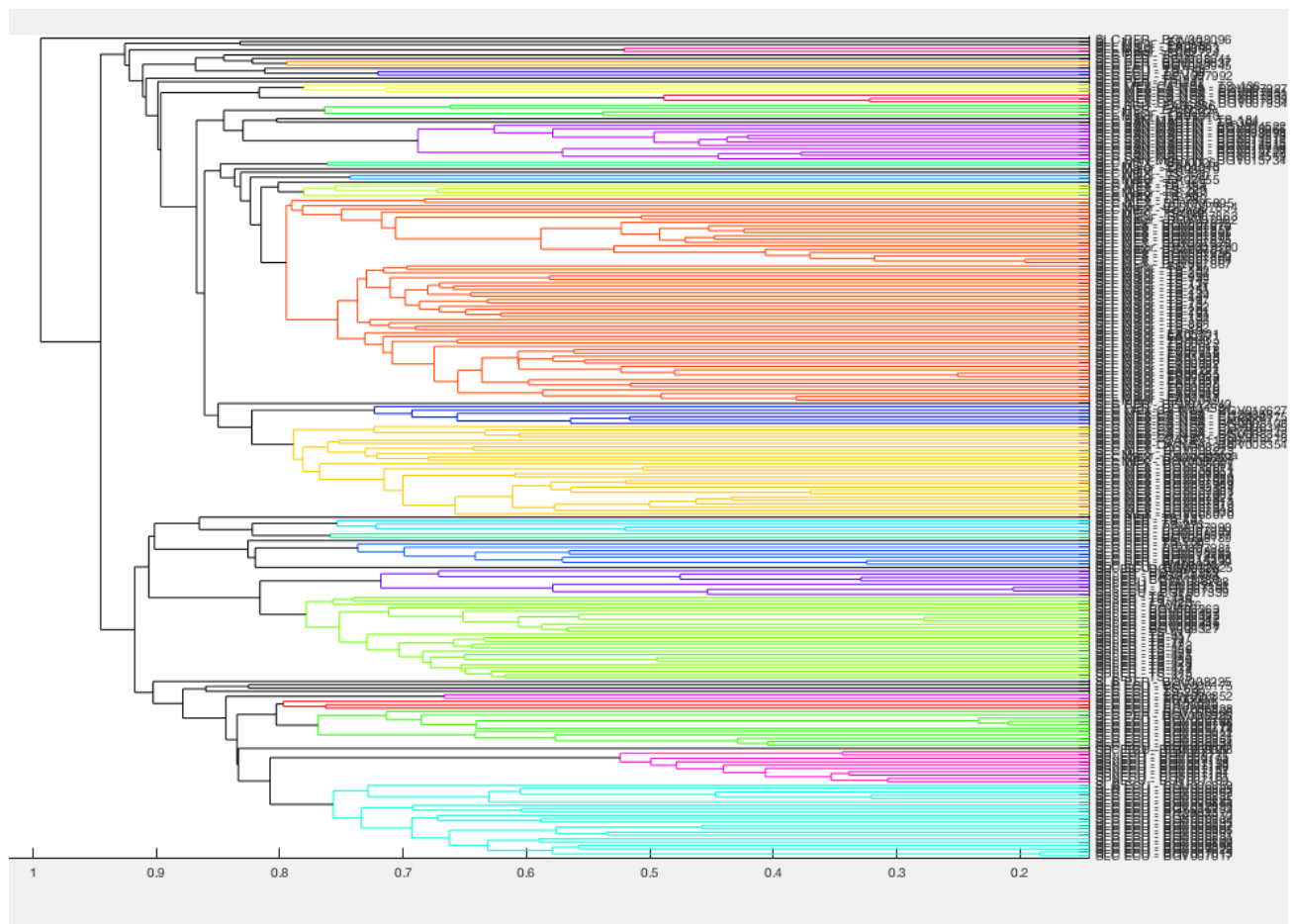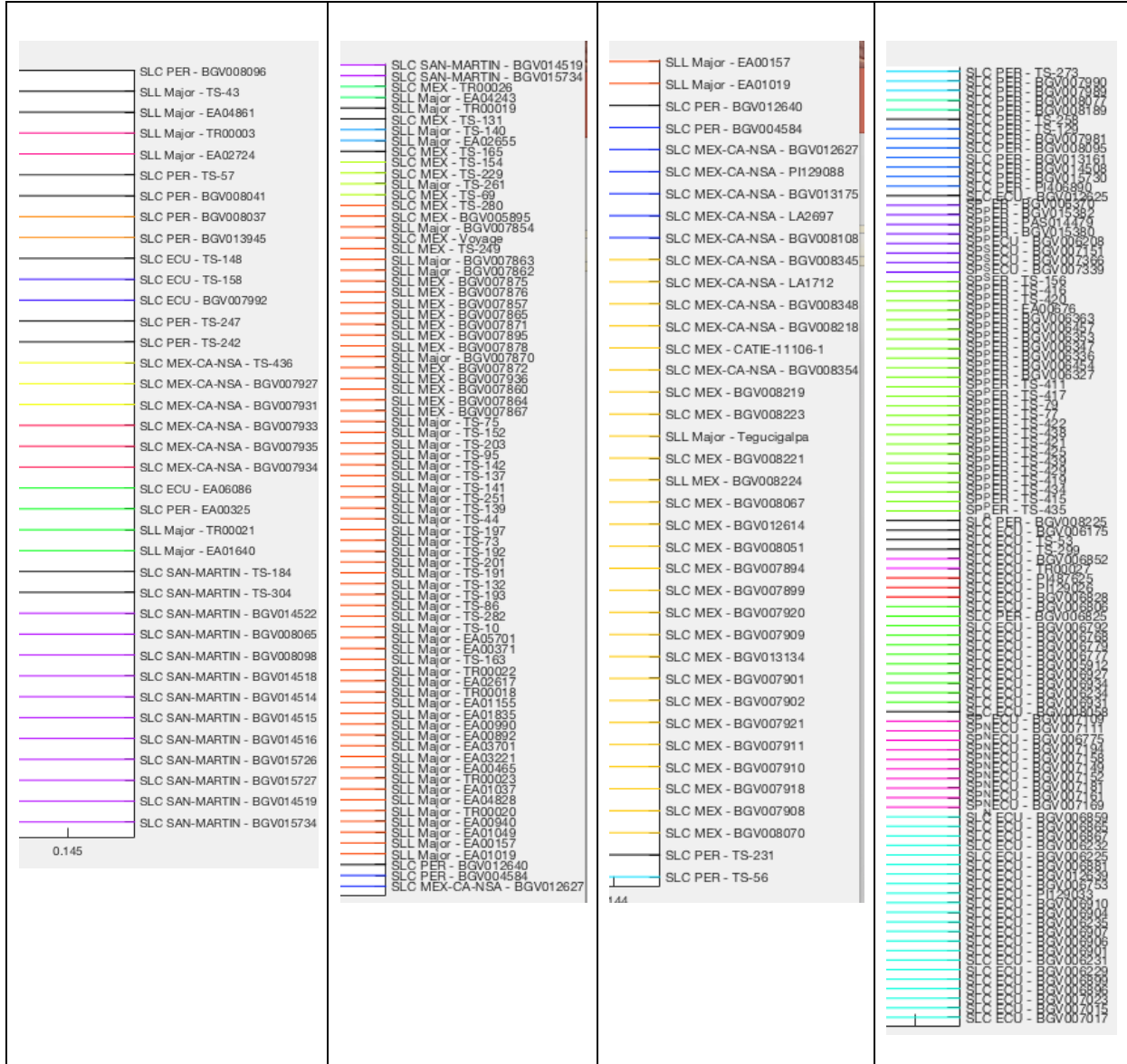USA
E-mail: EsthervanderKnaap@uga.edu

3

6 Appendix

Fig. 5: SV-based phylogenetic tree of the 245 tomato varieties.

Table 5: Zoom-in of the phylogenetic tree from Fig. 4: from left to right, we list the four regions from top-down of the tree.

The Shortest Way Travelling

Bogdan-George Gros, Emanuel Andrei Lestyan, Ionel Alexandru Put, Estera Groza

Abstract

The goal of this project is to present a new way of unique mapping and derationing for various businesses and ordinary people alike, using different shortest way algorithms. This project took birth from our curiosity in discovering if usual shortest way algorithms could be used to compete with large scale busines"s algorithms in shortest path finding. The application we realized is similar in usability with google platform, only its functionality was altered by with our own algorithms. The mobility of this app is somewhat limited at the moment as it"s restricted by the Google Play Store on certain devices. The distance between two points it"s computed with conventional Google Maps algorithm, although the shortest distance between multiple such points it"s computed with an algorithm with our own implementation. Examples of such algorithms include conventional methods including Bellman Kalaba, Ford, Dijkstra and Nearest Neighbor. We predict that such applications will be used on large scales in the future as mobility and speed are paramount in our everyday lives.

1 Introduction

In this project we propose a new way of unique mapping and derationing for various businesses and ordinary people alike. The focus of the project was to use different shortest way algorithms. This project took birth from our curiosity in discovering if usual shortest way algorithms could be used to compete with large scale business"s algorithms in shortest path finding. The application we realized is similar in usability with google platform, only its functionality was altered by with our own algorithms. The mobility of this app is somewhat limited at the moment as it"s restricted by the Google Play Store on certain devices. The distance between two points it"s computed with conventional Google Maps algorithm, although the shortest distance between multiple such points it"s computed with an algorithm with our own implementation. Examples of such algorithms include conventional methods including Bellman Kalaba, Ford, Dijkstra and Nearest Neighbor. We predict that such applications will be used on large scales in the future as mobility and speed are paramount in our everyday lives.

2 Related works

A number of applications similar to ours include, Waze, Here WeGo, Google Maps, Maps.me Google Maps. This application is provided by Google for web mapping purposes offering satellite imagery, aerial photography, street maps, 360 panoramic views of streets, real-time traffic conditions and route planning, aerial traffic and public transportation [2].

Waze, now in the ownership of Google, provides step-by-step browsing information, travel times and route details provided by users while downloading location dependent information over a mobile network. Waze describes the application as a community-based GPS navigation application that is free to download and use [4].

HERE WeGo is a web mapping and navigation service operated by AER technologies. Originally developed by Nokia as HERE Maps, the mapping software application was launched for Windows Phone and World Wide Web in 2013 as a revised version of Nokia Maps. HERE Maps was then launched for the Android platform on December 10 2014 and later for iOS on March 2015, and the name was changed to HERE WeGo in July 2016 [6].

Maps.me is a mobile app that offers offline maps using OpenStreetMap data. In November 2014 it was acquired by the Mail.Ru Group and became the part of My.com trademark. In September 2015 the app was open-sourced [7].

All of these applications have their pros and cons. For example Waze is better suited for drivers, Google Maps its better suited for pedestrians and cyclists, meanwhile Maps.me offers offline maps for poor or no internet zones.

2.1 The description of the algorithms

2.1.1 Dijkstra's algorithm

Dijkstra's algorithm (or the SPF algorithm of Dijkstra; Shortest Path First) is an algorithm for finding the shortest paths between nodes in a graph, which can be for example road networks. It was designed by computer scientist Edsger W. Dijkstra in 1956 and published three years later.

The algorithm exists in several variants; Dijkstra's original variant has found the shortest path between two nodes, but a more common variant pinches a node as a source node and finds the shortest paths from source to all the other nodes in the graph, producing a tree with the shortest way [9].

For a source node given in the graph, the algorithm finds the shortest path between this node and the other. Can also be used to find the shortest paths from a single node in a single destination node by stopping the algorithm after determining the shortest path to the destination node. For example, if chart nodes represent cities, and edge route costs represent driving distances between city pairs connected by a direct route (ignoring red lights, stop signs, toll roads, and other obstacles), Dijkstra's algorithm can be used to find the shortest route between the city and all other cities [10].

A large-scale application of the shortest path algorithm is network routing protocols, especially IS-IS (Intermediate System to Intermediate System) and Open Shortest Path First (OSPF). It is also used as a subroutine in other algorithms such as Johnson's.

The detailed steps used in Dijkstra's algorithm to find the shortest path from a single source vertex to all the other nodes in the given graph, are described in the following rows [11]:

1) Create a sptSet set (the shortest set of track trees) that tracks the spikes included in the shortest spindle, that is, whose minimum distance from the source is calculated and finalized. Initially, this set is empty.

2) Assign a distance value for all nodes in the input graph. Initialize all distance values as INFINITE. Assign the value of the distance as 0 to the tip of the source so it is selected first.

3) While sptSet does not include all spikes:
   a. Choose a u vertex that does not exist in sptSet and has a minimum distance value.
   b. Include u to sptSet.
   c. Update the value of the distance of all adjacent nodes of u. To update the distance values, iterate through all of the adjacent peaks. For each adjacent peak v, if the

sum of distance value of u and the weight of edge u-v is less than the distance value of v, then update the value of the distance v.

2.1.2 Bellman-Ford algorithm

The Ford-Fulkerson method as known as the Ford-Fulkerson algorithm (FFA) is a greedy algorithm that computes maximal flow of a flow network. In some cases called a "method" instead of an "algorithm" because the approach to find augmentation paths inside a residual plot is not wholly specified or specified into several implementations that have different operating times. It was published in the year 1956 by L. R. Ford, Jr. and D. R. Fulkerson. The name "Ford-Fulkerson" is often times used for the Edmonds-Karp algorithm, which is a fully-defined implementation of the Ford-Fulkerson method. The essence of the algorithm is as follows: so long as there is a path from the source (start node) to the finish (end node), with the available capacity on all the edges of the path, we send the flow along one of the paths. Then we find another way and so on. An available capacity path is called a magnification path [12].

Algorithm step-by-step [13]:

Input: Graph and a source src vertex

Exit: The shortest distance leading to all the nodes starting from src. If there exists a negative weight cycle, then the shortest time, the distances are not computed and the negative weight cycle is reported.

1) This step initiates the distances from source to all nodes as infinite, and the distance from the source itself is 0. Creating an array list[] of size | V | with all its values as infinite, except for dist [src] where src is the point of origin.

2) This step computes the shortest distances. Follow | V | -1 times where V | is the number of peaks in the given graph. Continue after each u-v edge:

If you dist [v]> dist + u, then update dist [v]

dist [v] = dist [u] + UV edge weight

3) This stage reports whether there is a negative weight cycle in the graph. The following is done for each u-v edge:

If dist [v]> dist + uv weight, then "The graph has a cycle of negative weight"

The last step shows that in the second step shorter distances are provided if the graph does not contain a negative weight cycle. If we repeat all the margins once and we get a shorter path for any peak, then there is a negative weight cycle [13].

The shortest paths are calculated in an ascending fashion. First, calculate the shortest distances that have at most one edge in the path. Then calculate the shortest paths with at most 2 margins and so on. After the third iteration of the outer loop, calculate the shortest paths with the margins at most. May be maximum - 1 edge in any simple way so that the outer loop is running v | - 1 times. The idea is that, assuming there is no loop loop, if we calculate the shortest paths with maximum margins, then an iteration on all edges guarantees the shortest path with no more (i + 1) margins [13].

2.1.3 The nearest neighbour algorithm

The nearest neighbour algorithm was one of the first algorithms used to solve the problem of the travelling salesman. In it, the vendor begins in a random city and repeatedly visits the nearest town until all were visited. Quickly performs a short tour, but usually is not the optimal one.

The algorithm is easy to implement and execute quickly, but sometimes it may miss shorter routes that are easy to see with a human understanding due to its "greedy" nature [1].

As a general guide, if the final stages of the tour are comparable in length to the first stages, then the tour is reasonable; if they are much larger then it is likely that there are much better trips. Another check is to use an algorithm such as the lower limit algorithm to estimate whether this tour is good enough [3].

In the worst case scenario, the algorithm results in a tour that is much longer than the optimum tour. To be exact, for each constant r there is an instance of the vendor's problem that travels so that the duration of the tour calculated by the nearest neighbour algorithm is greater than r or the optimum tour length. Moreover, for each number of cities there is a distribution of the distances between the cities for which the neighbour's closest heurist produces the worst possible tour. (If the algorithm is applied to each peak as the starting point, the best path found will be better than at least N / 2-1 other runs, where N is the number of peaks) [5].

The nearest neighbour algorithm may not find a feasible tour at all, even when there is one.

3 Proposed Solutions

This project took birth from our curiosity in discovering if usual shortest way algorithms could be used to compete with large scale business‟s algorithms in shortest path finding. The application we realized is similar in usability with google platform, only its functionality was altered by with our own algorithms. The mobility of this app is somewhat limited at the moment as it‟s restricted by the Google Play Store on certain devices.

3.1 Android Studio

This application was developed in Android Studio the official integrated development environment (IDE) for Google's Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development.

One of the best platforms for realizing android applications, beneficiating from: Gradle-based build support; Android-specific refactoring and quick fixes; Lint tools to catch performance, usability, version compatibility and other problems; ProGuard integration and app-signing capabilities; Template-based wizards to create common Android designs and components; A rich layout editor that allows users to drag-and-drop UI components, option to preview layouts on multiple screen configurations; Support for building Android Wear apps; Built-in support for Google Cloud Platform, enabling integration with Firebase Cloud Messaging (Earlier 'Google Cloud Messaging') and Google App Engine; Android Virtual Device (Emulator) to run and debug apps; It is one of the best choices for those looking to begin learning or seasoned developers.

3.2 Mobile Devices

As of last decades mobile devices began booming on the market as products, be it their mobility, accessibility, price, or whatever other reason almost everyone owns a mobile device. Mobile devices became as much a distraction as a necessity in our everyday lives, now more than ever when communication and boredom are at an all-time high, mobile devices seem to continue their upward trend in popularity among all ages and places.

And although this miracle of modern engineering can be carried in our pocket, making it work is no easy task, be it hardware or software.

Large software apps are developed by teams of hundreds or even more people, and for good reasons, games and small apps may be able to be created with sheer dedication, but a bigger one is just not feasible to do alone, that is not to say that lone people didn‟t create amazing apps, most of google play store is full of those, but the resources, tasks and deadlines can be discouraging.

3.3 Google maps API

The Google Maps API is used for the most part of the project, as we do not possess the necessary info to create such an ambitious data dependant project most of the information is provided by it. Everything from the actual map to directions is parsed through calls and responses from the API server. As such an internet access is imperative at this stage. With time this dependency could be avoided by including a cache format, databases and other data storing utilities [8].

3.4 Implementation

3.4.1 Design

The design of this application is fairly straightforward considering the map is generated by Google, the only major design changes could be buttons, edit texts, plain texts for informing the user, data introduction and so on. The app admittedly is not very stylish but with a sleek look, and functionality, it stands its own ground.

As for implementation, the algorithms, do what they are designed to do the best, find the shortest path between two points. Regarding the distance between the points and the course taken, Google has the data no one could ever dream of, gathered over a large span of time, it is as reliable as anything one can get their hands on. As for what our project brings new to the table: a reliable way to determine the most efficient way you can get through multiple points on a map. This would massively benefit cargo and transport businesses as the cost of transporting goods would be reduce and such, profits maximized. As for future previsions, we would predict that applications like these will be just as or more popular than today. As transport, speed and efficiency become more and more pronounced in our everyday lives, without a doubt programs similar to our project will proliferate and become more and more refined.

3.4.2 Program Code

As simplistic as the idea of interconnecting dots on a map sounds, as laborious the implementation can be. Fortunately the Google API does make the lives of developers and users alike much easier. Through their continuously improving platform they are able to deliver functions and features helping both aspiring and senior programmers.

As for the algorithm, first ensuring a stable connection to Google API is paramount, for initializing the map, retrieving coordinates and so on.

After the process was successfully completed, then come the origin and destination markers, after those were chosen their data must be parsed through the API, for processing and getting the distance, available routes and so on.

```
public void onMapReady(GoogleMap googleMap) {
        mMap = googleMap;
        Log.d("mylog", "Added Markers");
        mMap.getUiSettings().setZoomControlsEnabled(true);
        //detect if acces to fine location is available
        if (ActivityCompat.checkSelfPermission(this, Manifest.permission.ACCESS_FINE_LOCATION) !=
PackageManager.PERMISSION_GRANTED && ActivityCompat.checkSelfPermission(this,
Manifest.permission.ACCESS_COARSE_LOCATION) != PackageManager.PERMISSION_GRANTED) {
            ActivityCompat.requestPermissions(this,                                         new
String[]{Manifest.permission.ACCESS_FINE_LOCATION},LOCATION_REQUEST);
            return;
        }
        mMap.setMyLocationEnabled(true);
        mMap.setOnMapLongClickListener(new GoogleMap.OnMapLongClickListener() {
          @Override
          public void onMapLongClick(LatLng latLng) {
             //Reset marker when already 2
```

```
            if (listPoints.size() == 10) {
               listPoints.clear();
               mMap.clear();
            }
            //Save first point select
            //Create marker
            MarkerOptions markerOptions = new MarkerOptions();
            markerOptions.position(latLng);
            listPoints.add(markerOptions);

            if (listPoints.size() == 1) {
               //Add first marker to the map

markerOptions.icon(BitmapDescriptorFactory.defaultMarker(BitmapDescriptorFactory.HUE_GREEN));
            } else {
               //Add second marker to the map

markerOptions.icon(BitmapDescriptorFactory.defaultMarker(BitmapDescriptorFactory.HUE_RED));
            }
            mMap.addMarker(markerOptions);
         }
      });
   }
```
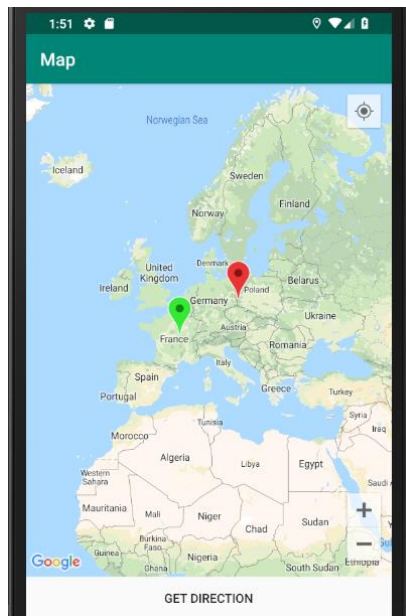

Fig.1 –Setting the destination

The data must be parsed through an URL that the google API can recognize, this very URL can be used in a normal browser window directly accessing google maps.

```
     private String getUrl(LatLng origin, LatLng dest, String directionMode) {
        // Origin of route
        String str_origin = "origin=" + origin.latitude + "," + origin.longitude;
        // Destination of route
        String str_dest = "destination=" + dest.latitude + "," + dest.longitude;
        // Mode
        String mode = "mode=" + directionMode;
        // Building the parameters to the web service
        String parameters = str_origin + "&" + str_dest + "&" + mode;
        // Output format
        String output = "json";
```

```
        // Building the url to the web service
        String url = "https://maps.googleapis.com/maps/api/directions/" + output + "?" + parameters +
"&key=" + getString(R.string.google_maps_key);
        return url;
```

Once we have this data the algorithm compares the distance from the origin to each destination, the smallest is chosen, then that marker becomes the origin, the first origin is noted as visited and the process repeats again for the second, choosing the smallest distance and so on. This process continues until the algorithm reaches the last destination, point at which it stops and computes through a line the most efficient route between nodes. But that doesn't always happen as we saw in an earlier paragraph nearest neighbour doesn't have a 100% shortest path finding rate, so as to mitigate that we use other algorithms like Bellman, Ford or Dijkstra to compute the shortest way.

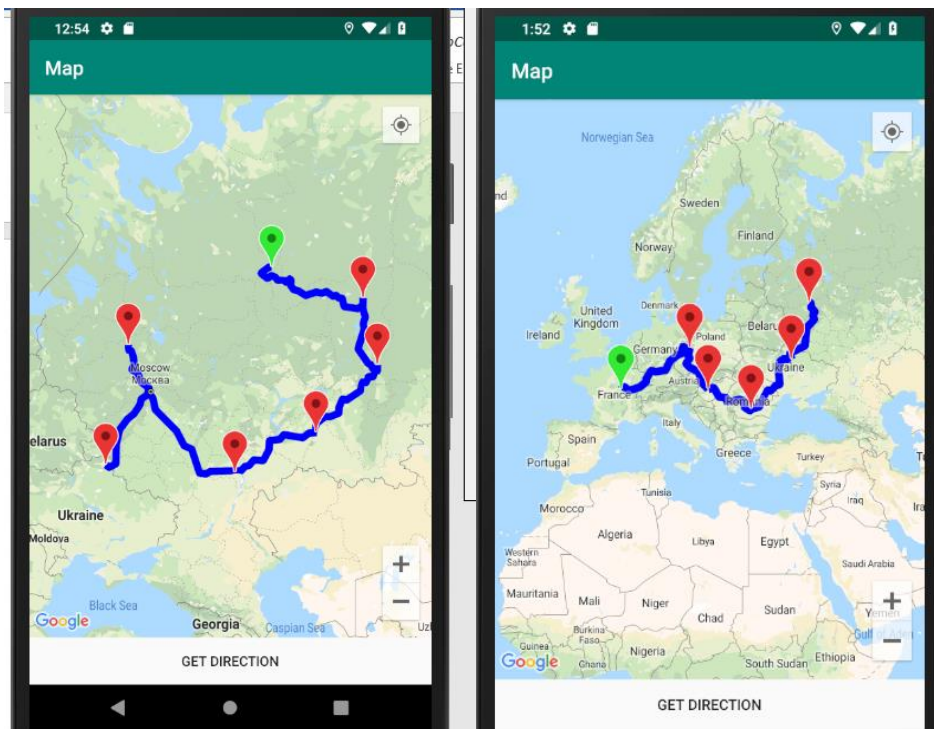Fig.2 – Application printscreen

```
class Graph      {
    class Edge
    {
        public int src, dest, weight;
        public Edge()
        {
            src = dest = weight = 0;
        }
    };
    int V, E;
    public Edge[] edge;
    public Graph(int v, int e)
    {
        V = v;
        E = e;
        edge = new Edge[E];
        for (int i = 0; i < e; ++i)
            edge[i] = new Edge();
    }

public void BellmanFord(Graph graph, int src, int[] dist) {
    int V = graph.V, E = graph.E;
    for (int i = 0; i < V; ++i)
        dist[i] = Integer.MAX_VALUE;
    dist[src] = 0;
    for (int i = 1; i < V; ++i)
    {
        for (int j = 0; j < E; ++j){
            int u = graph.edge[j].src;
            int v = graph.edge[j].dest;
            int weight = graph.edge[j].weight;
            if  (dist[u]   !=   Integer.MAX_VALUE   &&
dist[u] + weight < dist[v])
                dist[v] = dist[u] + weight;
        }
    }
}
```

After the route has been checked polylines between markers will appear indicating the shortest way found.

Between these bits of code are classes made for parsing JSON and manipulating the URL, which only serve the purpose of „bridges" between our app and the google API.

4 Conclusions

As we all know trying to develop an app, especially one that someone else already did better is no easy task, but from this experience we all learned something; be it to persevere, to learn to push forward when nothing works, to code in Java, use Android Studio and this API, or by simply being part of this, we all feel that we moved forward a bit.

References

[1] G. Gutin, A. Yeo and A. Zverovich, *Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP*. Discrete Applied Mathematics 117 (2002).
[2] ***, https://en.wikipedia.org/wiki/Google_Maps
[3] Moore, Edward F. (1959). *The shortest path through a maze*. Proc. Internat. Sympos. Switching Theory 1957, Part II. Cambridge, Massachusetts: Harvard Univ.
[4] ***, https://en.wikipedia.org/wiki/Waze
[5] Yen, Jin Y. (1970). "*An algorithm for finding shortest routes from all source nodes to a given destination in general networks*". Quarterly of Applied Mathematics.
[6] ***, https://en.wikipedia.org/wiki/HERE_WeGo
[7] ***, https://en.wikipedia.org/wiki/Maps.me
[8] "*What is the Google Maps API?*" Archived from the original on April 20, 2012
[9] ***, https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm
[10] Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2001). "Section 24.3: *Dijkstra's algorithm*". Introduction to Algorithms (Second Ed.). MIT Press and McGraw–Hill.
[11] Knuth, D.E. (1977). "*A Generalization of Dijkstra's Algorithm*". Information Processing Letters
[12] ***, https://en.wikipedia.org/wiki/Ford%E2%80%93Fulkerson_algorithm
[13] ***, https://www.geeksforgeeks.org/bellman-ford-algorithm-dp-23/

Bogdan-George GROS
"Aurel Vlaicu" University of Arad
Applied Informatics
Str. Elena Drăgoi, nr. 2, 310330,
ROMANIA
E-mail: gros.bogdan@yahoo.com

Emanuel Andrei LESTYAN
"Aurel Vlaicu" University of Arad
Informatics
Str. Elena Drăgoi, nr. 2, 310330,
ROMANIA
E-mail: lestyanemanuel@gmail.com

Ionel Alexandru PUT
"Aurel Vlaicu" University of Arad
Informatics
Str. Elena Drăgoi, nr. 2, 310330
ROMANIA
E-mail: ionelput98@gmail.com

Estera GROZA
"Aurel Vlaicu" University of Arad
Informatics
Str. Elena Drăgoi, nr. 2, 310330
ROMANIA
E-mail: estygroza@yahoo.com

.NET Obfuscator

Svetoslav Hadzhiivanov

Abstract

This paper examines the importance of .NET Obfuscators in our digital society, the techniques used to protect our applications and compares existing obfuscation solutions. The author of the paper is proposing a new obfuscation tool which is better than others because its features are made in a different way.

1 Introduction

.NET takes a large percentage of all commercial and non-commercial Desktop applications. It is based on .NET Framework. There are two types of code: Native, also known as unmanaged (e.g. C/C++) and Managed (.NET and Java) [1]. The native code is compiled to work directly with the OS whereas the managed code is precompiled and then processed by JIT (Just-in-time Compiler) [2] (fig. 1). Before static or dynamic compilation to machine code, the .NET Framework uses Intermediate language (IL) [3]. The IL code is easily understandable by .NET decompilers (like dnSpy) and anyone with minimal programming skills can read or modify it unlike the native code. Therefore, .NET Obfuscator's role comes into play.
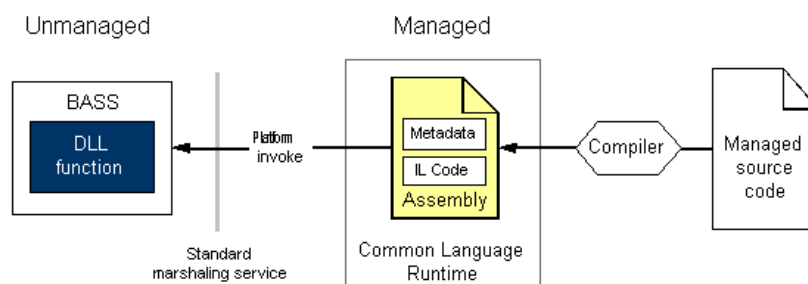

Fig. 1 Native vs Managed code

.NET Obfuscators are trying to reorganize the code into a less understandable one by Reverse Engineering enthusiasts or professionals. A good reverser will always be able to reverse („crack") application, even though it might be heavily obfuscated. The point of obfuscating applications is to slow down reversers as much as possible or even better to make them give up in the reversing process because it would make them spend a lot of time to follow the obfuscated code deeper and

understand it. The most protected applications are usually Client/Server based with encrypted network traffic and heavily obfuscated client-side code because they rely on their servers whose code is inaccessible unless there is a vulnerability for example in their API. However, it's not impossible to crack such applications because messages send from the server and received by the client can be changed. Basically, one unobfuscated application is nearly as same as providing its source code.

2 .NET Obfuscator features explained

To compare a few products, their specifications must be known. Only the most important obfuscator features will be discussed below.

2.1 String Encryption

It encrypts each string with an algorithm using a key and it is later decrypted at runtime (after application's execution). This feature is vital because it might lead to the code's nature. In the figure below, the first argument represents the encrypted string and the second one is the encryption key. However, additional arguments can be added to fool reversers.


```
return <<EMPTY_NAME>>.TokenizerStream("JR04zEejNAtwta1xtIctg7TqL24WpVtVhdXtAb5FY7xyiEkXRFWVwUE/
  H8C1EAoSZ1CHBtXej3QRx7YFBSaI4GX8Xjiojs2UrWhgCII7rdOEi8jO0MR1mEhmr8ZZN6Vn", "eqvIjuPoVeRlPPWPNkME");
```

Fig. 2 String Encryption

2.2 Anti Tampering

It makes the code harder to be modified. In the figure below, it looks like there are two empty functions but they are really not. It's kinda obvious because ReadStub function returns a byte array and there is not a such code in the function. If there is an importance level between 1 and 3, this feature would be rated as 2.


```
internal class Program
{
    // Token: 0x06000004 RID: 4 RVA: 0x00002018 File Offset: 0x00000418
    [STAThread]
    private static void Main()
    {
    }

    // Token: 0x06000005 RID: 5 RVA: 0x00002164 File Offset: 0x00000564
    private static byte[] ReadStub(string key)
    {
    }
```

Fig. 3 Anti Tampering

2.3 Control Flow Obfuscation

The Control Flow Obfuscation adds additional instructions as shown in the figure below. The additional instructions are additional code added between the real instructions which often includes fake calls to fool reversers.

```
int num = -123892 + 42017 + 81875;
do
{
    if (num == -3070 + 89754 - 86684)
    {
        num = 30929 + 20617 - 51545;
    }
}
while (num != 38534 + 21367 - 59900);
```

Fig. 4 Control Flow Obfuscation

2.4 Invalid Metadata

This feature adds invalid metadata to the modules in order to prevent methods from being decompiled.


```
[STAThread]
private static void Main()
{
    /*
n exception occurred when decompiling this method (06000001)

CSharpCode.Decompiler.DecompilerException: Error decompiling System.Void Loader.Program::Main()
---> System.OverflowException: Arithmetic operation resulted in an overflow.
  at ICSharpCode.Decompiler.ILAst.ILAstBuilder.StackSlot.ModifyStack(StackSlot[] stack, Int32 popCount, Int32 pushCount
    \ICSharpCode.Decompiler\ILAst\ILAstBuilder.cs:line 47
  at ICSharpCode.Decompiler.ILAst.ILAstBuilder.StackAnalysis(MethodDef methodDef) in C:\projects\dnspy\Extensions\ILSp
  at ICSharpCode.Decompiler.ILAst.ILAstBuilder.Build(MethodDef methodDef, Boolean optimize, DecompilerContext context)
    \ILAstBuilder.cs:line 269
  at ICSharpCode.Decompiler.Ast.AstMethodBodyBuilder.CreateMethodBody(IEnumerable`1 parameters, MethodDebugInfoBuilder
    \ICSharpCode.Decompiler\Ast\AstMethodBodyBuilder.cs:line 112
  at ICSharpCode.Decompiler.Ast.AstMethodBodyBuilder.CreateMethodBody(MethodDef methodDef, DecompilerContext context,
    StringBuilder sb, MethodDebugInfoBuilder& stmtsBuilder) in C:\projects\dnspy\Extensions\ILSpy.Decompiler\ICSharpCo
--- End of inner exception stack trace ---
  at ICSharpCode.Decompiler.Ast.AstMethodBodyBuilder.CreateMethodBody(MethodDef methodDef, DecompilerContext context,
    StringBuilder sb, MethodDebugInfoBuilder& stmtsBuilder) in C:\projects\dnspy\Extensions\ILSpy.Decompiler\ICSharpCo
  at ICSharpCode.Decompiler.Ast.AstBuilder.<>c__DisplayClass89_1.<AddMethodBody>b__0() in C:\projects\dnspy\Extensions
*/;
    }
```

Fig. 5 Invalid Metadata

2.5 Symbols Renaming

This feature simply renames: namespaces, methods, parameters, fields, etc.


```
[STAThread]
private static void INVOCATION_FLAGS()
{
```

Fig. 6 Symbols Renaming

3   Existing Obfuscation Solutions Overview

The best obfuscator is not only the one being able to protect an application strongly but it should also keep the performance of the application as fast as possible. For example, there might be a well-obfuscated application which is starting for 45 seconds. If it's loading so slowly, nobody would like it.

Another thing that we must have in mind is the obfuscated file. The output file size should also be kept to minimum. A disadvantage of .NET is that the assembly could easily be corrupted unlike in a native language. For example, calling an API in .NET with wrong arguments can

throw an exception while in C/C++ it will only return a certain error code and will continue the code execution.

3.1 .NET Reactor

.NET Reactor is advertised as a powerful code protection and software licensing system tool. It is one of the most preferred obfuscation tools by people. .NET Reactor's only advantage is the native loader which makes the obfuscated application look like it is not .NET application and some inexperienced people might be fooled. It doesn't corrupt files which is good but its main disadvantage is that it is deobfuscatable by de4dot (.NET deobfuscator and unpacker) which makes it useless in production.

.NET Reactor's main features are:

- Native Loader and Native Code Generation
- Control Flow Obfuscation
- String Encryption
- Anti Tampering
- Resource Encryption and Compression
- Dependency Merging

3.2 Eazfuscator.NET

Eazfuscator.NET is probably one of the best obfuscators used by many people. The only feature that deserves attention is their code virtualization. The code virtualization replaces developer's code with a certain byte sequence which is only understandable by their virtual machine or at least it is supposed to be so.

We should note that the application cannot virtualize all methods. Because of performance and file size issues. Same as in .NET Reactor, Eazfuscastor.NET does not break applications but it is also supported by de4dot which makes the final decision.

Eazfuscator.NET's main features are:

- Code and data virtualization
- Control Flow Obfuscation
- String Encryption and Compression

- Symbols renaming (including XAML renaming)
- Resource Encryption and Compression
- Assemblies merging and embedding
- Automatic code optimization

3.3 NETGuard

NETGuard is being updated frequently which makes the code which it produces harder for deobfuscation. Currently, de4dot doesn't support it. It also has a native loader like .NET Reactor and a proven to work Anti Dump protection unlike other not working solutions in other products. It is vital to prevent the executable from being dumped from memory because even the best obfuscation can be neutralized by dumping.

NETGuard's main features are:

- Native Loader
- String Encryption and Compression
- Symbols renaming (including XAML renaming)
- Resource Encryption and Compression
- Anti Dump
- Anti Tampering
- Anti De4dot
- Anti Debug
- Variable Melting
- Reference Proxy
- Invalid Metadata

3.4 SmartAssembly
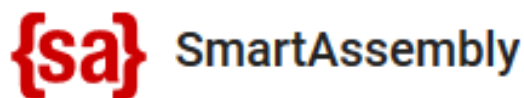
SmartAssembly is the weakest and easily reversable obfuscator. It is often used by malware campaigns.

SmartAssembly's main features are:

- Symbols renaming
- String Encryption
- Control Flow Obfuscation
- Reference Proxy
- Anti Tampering
- Dependency merging and compression

3.5 Obfuscator's Review Conclusion

Comparing all the features, we can say that NETGuard is the best commercial obfuscator from the discussed above. One of its strongest sides are its Anti Dump and the frequent updates.

We can give the second place to Eazfuscator.NET due to its code and data virtualization.

In addition to all mentioned obfuscator features, we must add the junk code obfuscation. Junk code is basically code doing nothing but it helps a lot in the process of obfuscation.

4   Proposed Obfuscator Overview

This obfuscator is better than the reviewed above because its features are made in different way which makes it a new challenge for deobfuscators and of course Reverse Engineers. It is easy to use and multithreaded. Multithreaded means whenever a long-time operation button is pressed, it will not block the user interface thread (main thread of the application). The user has to only drag and drop

Features:

- Code and data virtualization
- Symbols renaming
- String Encryption
- Control Flow Obfuscation
- Anti Tampering
- Anti Dump
- Anti Debug
- Anti De4dot
- Constants Melting
- Reference Proxy
- Invalid Metadata

There are situations where a part of a program should be very well protected (e.g. banking software) and in this case, symbols and strings protection is just not enough. The code virtualization feature is changing the way the program is presented in memory and on disk. It translates the .NET byte code to a totally unrecognizable random-generated byte sequence, which still perfectly works at runtime.

In fig. 2, a string encryption image is presented. However, these two arguments can be easily recognized by more experienced people. The first one is the encrypted string and the second one is the key. Paper's Obfuscator is adding additional arguments to each encrypted string to try to make it look like a random function call.

The Control Flow Obfuscation is improved by adding additional loop instructions.

Anti Debug and Anti De4dot features are pretty self-explanatory and Anti Dump is preventing application from being dumped in memory.

Constants Melting feature is basically moving constants to new methods which are returning them in an encrypted way.

Reference Proxy feature was not described above but however, it is trying to encode and hide calls to types/methods/fields.

5 Conclusion

This paper presents the most important obfuscation techniques and compares some of the most famous obfuscators. In the future, paper's example obfuscator will be updated with a GUI. There is one more important thing that products should not be judged by their GUI but how effective they are.

Acknowledgement: This work was supervised by Prof. *Katalina Grigorova* and Dr. *Metodi Dimitrov* from "Angel Kanchev" *University of Ruse*.

References

[1] Differences between native and managed code, https://stackoverflow.com/questions/855756/difference-between-native-and-managed-code
[2] Just-in-time Compiler, https://aboullaite.me/understanding-jit-compiler-just-in-time-compiler/
[3] Intermediate language, https://www.techopedia.com/definition/24290/intermediate-language-il-net

Svetoslav HADZIIVANOV
University of Ruse
Department of Informatics and
        Information Technologies
8, Studentska str., Ruse 7017
BULGARIA
E-mail: warrolen@gmail.com

RUBIKon

Horia Iuga, Andreas Coş

Abstract

The "RUBIKon" project aims to build a robot to solve the Rubik cube as soon as possible, combining robotic knowledge with software engineering. Cube analysis is done using a video camera attached to a Raspberry Pi 3 card, which will then take the images and process them, extract the colors with the recognition software, and build a virtual image of the mixed cube. It will calculate the possible solutions for 3x3x3 cube solving and the information of the 3 engines will be transmitted. The engines will move the physical mechanism made of Lego Tehnic parts.

1 Introduction

In the last decade, robotics education has flourished in Romania, initiatives have been developed as an expression of ideas promoted by "constructivist learning", such as robotics workshops that help develop skills and knowledge in STEM (Science-Technology-Engineering-Mathematics). Robots are technological artifacts that allow students to experiment, produce, collect data, process them, and gain concepts. Robotics is a transdisciplinary subject that allows the formation and development of the technical creativity and algorithmic thinking of the students. We chose as a project to build a robot called "RUBIKon" capable of solving any Rubik 3x3x3 cube in a short time. The Rubik Cube is a puzzle-based puzzle game invented in 1974 by the Hungarian architect sculptor and professor Ernő Rubik. The number of possible shuffles is equal to 8! x 37 x 12! x 210 $\approx$ 4.33 x 1019, but only one position is the correct one.

2 Description of the project

We assembled the robot by a plan made by us. This process took around three months. Then we started to learn Python and tried to program the robot. It was a very hard process because nobody published on the internet a similar program and there were few robots that could solve the rubik cube. After six months we had a working software of 3x3x3 cube. It had about 2200 lines of code but there were a lot of useless moves.

In present, the software has about 1200 lines of code and all useless moves were removed. We also improved the solving time from three minutes to around 90 seconds, depending on the mode selected (displayed in figure 3).

2.1 Hardware modules

For the hardware modules we chose:

- **One Raspberry Pi3 B**

We use it as a small computer running an adapted version of Linux. It's very tiny and has enough processing power for this project.

- **One Raspberry Pi camera V2**

We chose this camera because it's very easy to use with Raspberry Pi.

- **Three stepper motors Nema 17**

These motors have enough power to rotate all Lego mechanisms.

- **Three drivers A4988 for stepper motors attached to three expansion boards**

The drivers control the motors based of the Raspberry Pi indications.

- **Wires**

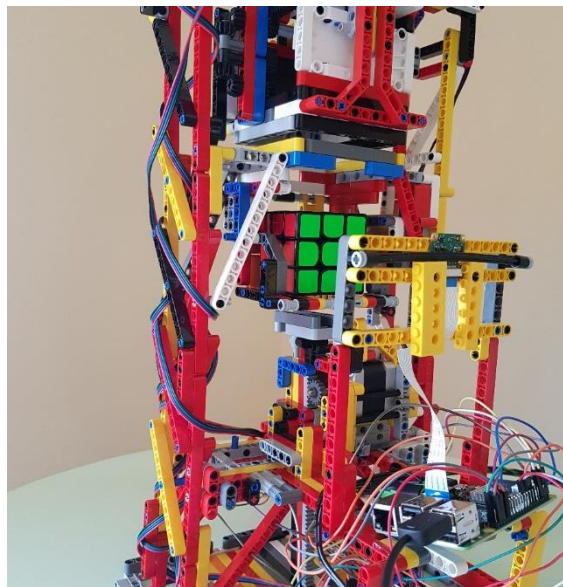- **One 24V 5A power supply**

- **Lego pieces**

Fig. 1: RUBIKon Robot-hardware

2.2 Software modules

The robot is exclusively programmed in Python 3.5. We designed an interface (see Fig. 1) based on Tkinter library which opens a window when the program runs (see Fig. 2).

```
class Window(object):

    def __init__(self, master):
        self.master = master
        self.master.title("Project RUBIKon")
        self.master.geometry("600x400+300+200")
        self.AddWidgets()
    def AddWidgets(self):
        self.TopFrame = Frame(self.master, bg = "#42f498", width = 600, height = 60)
        self.TopFrame.grid(column = 0, row = 1)
```

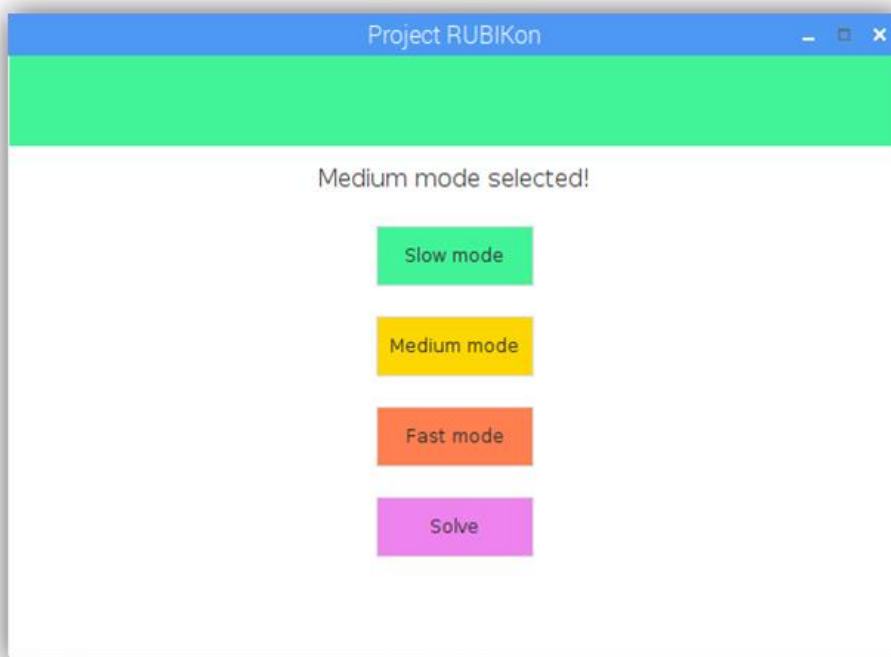Fig. 2: Tkinter interface – Example code

Fig.3: Tkinter Interface – Mode Selection

We can choose between three modes: **Slow, Medium and Fast**. Than the "**Solve**" button appears and when we click it, the solving procedures are initiated:

•   **Scan procedure:** At this state, the camera module is turned on. Then the motors rotate the cube and all six sides are photographed (see Fig. 4) and the images are stored in a folder.

```
camera.start_preview()
camera.capture('/home/pi/cubescan/rubiks-side-R.png')
elevatorUp()
```

Fig. 4: Part of the code from Scan procedure

- **Image processing:** For this step we use a python library called OpenCV. We apply different filters to all images (ex. Gaussian blur, Canny and more) to evidence the edges of the squares present in the 3x3x3 cube. Then we extract pixels (format is RGB) from all squares and the software creates a virtual image of the scrambled cube (see Fig.5). The state of the cube is sent to the solver who generates a solution.
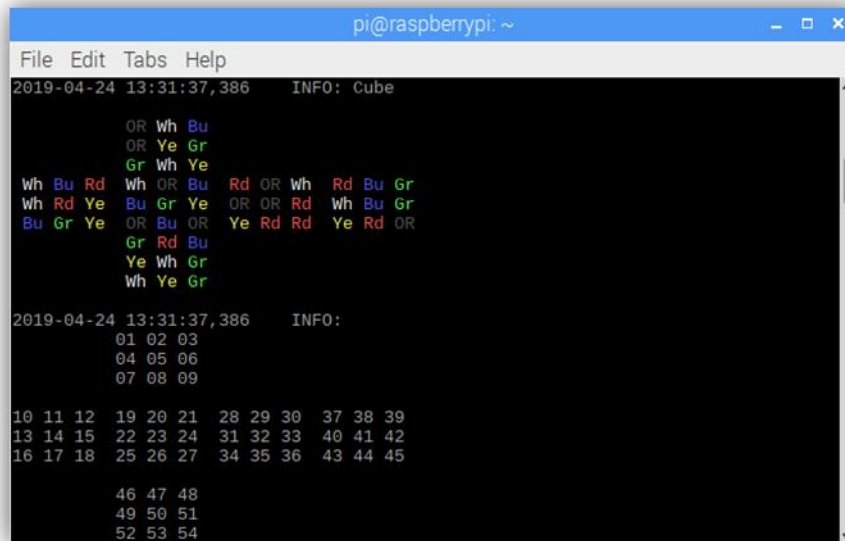

Fig. 5: Virtual image of the scrambled cube

- **Cube solution**: The solution is processed and adapted for the motors. At this step starts the physical solving of the cube. After the cube is solved, it's unlocked a new feature: "Custom pattern". (see Fig. 6)
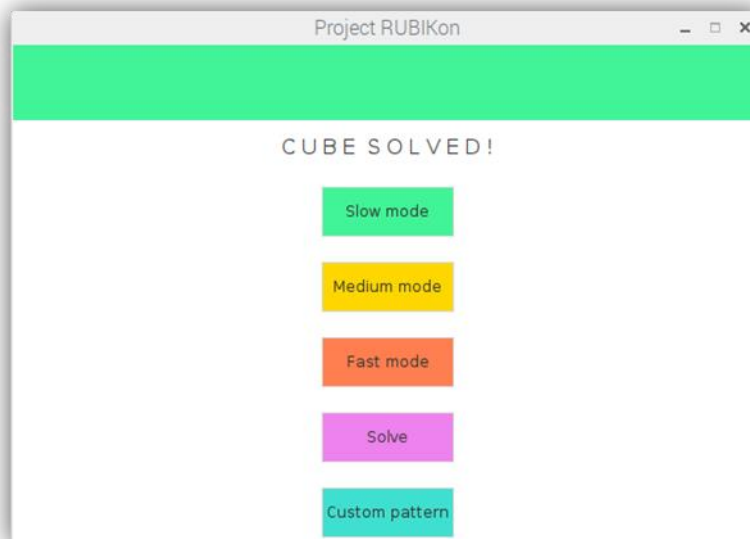

Fig. 6: Solved cube – Custom pattern

We can now touch the "Custom pattern" button that takes us to the next page.

Fig. 7: The Cube – pattern

From this page (see Fig. 7) we can choose from four different patterns: *Checkboard, Six Spots, Union Jack* (see Fig. 8) and Six Crosses.

Fig. 7:Union Jack pattern

3 Results

Currently, the 3x3x3 cube can be solved in about 90 seconds, including the scanning process. From our tests so far (Table 1), all 3x3x3 cubes have been successfully solved. However, sometimes a 3%

error occurs when colors are misidentified, due to the different light conditions present in demonstration room.

Nr. Cubes tested	Solved	Unsolved	
		Because of the lightning conditions	Other reasons
100	97	3	0

Table 1: RUBIKon tests 3x3x3

3   Conclusions and further research directions

The modeling of the RUBIKon robot involved the development of robot's algorithms, the loading of the written programs in the control units and their execution. This project has contributed to the formation and development of our personality, has developed our technical creativity, logical thinking and algorithmic thinking. The project called for the implementation of modeling skills, mechanical, electromechanical, electronic, optical, computer and algorithmic programming.
As a further development of the project, we remember improving the software to allow the robot to solve any cube in the range of 3x3x3 to 6x6x6. Theoretically, it could solve any cube (as the number of squares in width), but we limit it to 6x6x6 because with the increase in the number of squares, the cube size also increases and implicitly it will be necessary to approach another construction (from lego pieces) of the robot.

References

[1] Mueller Andreas C., Guido Sarah. *Introduction to Machine Learning with Python.* O'Really Media, 2016
[2] Stowell L., Dikins R., *Programare pentru începători folosind Python,* Editura Litera, București, 2018
[3] Python tutorials - http://learn-robotics.com/
[4] Python tutorials - https://www.oreilly.com/library/view/python-in-a/0596001886/ch04s03.html
[5] Python tutorial - https://www.python-course.eu/passing_arguments.php
[6] RaspberryPi tutorial - https://www.raspberrypi.org/documentation/

Horea IUGA
 "Horea, Closca si Crisan" High School
Alba Iulia
B-dul 1 Decembrie 1918,No. 11
ROMANIA
E-mail: iugahoria@gmail.com

Andreas COŞ
 "Horea, Closca si Crisan" High School Alba
Iulia
B-dul 1 Decembrie 1918,No. 11
ROMANIA
E-mail: andreascos2002@gmail.com

Escape the garden – 3D game for VR

Iskren Ivanov

Abstract

Nowadays, Virtual Reality (VR) is becoming more and more popular and it is often used for various purposes in science. Escape the garden is an Android application that uses VR glasses and VR controller. The current paper discusses various technologies, used in the application as well as some of its main features. Some aspects of VR technology have also been addressed.

1 Introduction

Over the last few years, there has been a big interest in VR games. In fact, the first VR glasses have been created long ago. Virtual reality has existed for even longer, but modern technology enables a much wider range of people to touch upon this creation of science.

Escape the garden is a single player 3D VR [1] game, which can be played on almost any Android phone. It is made with Unity [2], C# and Google VR SDK [3]. As the name suggests, the main goal for the player is to escape the garden where he is trapped. This can be done by solving a series of puzzles and finding the hidden key for the gate that leads out. The game is very intuitive and it is easily assimilable for everyone who gets involved with it. The fact that it uses only two buttons - one for movement and one for interacting with objects, confirms how comfortable it is. Rotating the camera in the game is achieved through the player head move (both rotating and tilting) as in all games of this type.

The game is meant to be played on Google Cardboard [4]. There are some newer technologies for VR like Oculus [5] and Vive [6], but they are not so accessible, because of their high price. However, Escape the garden can be played on these platforms too.

Besides being fun, it is scientifically proven that the games develop a lot of positive qualities in the player. In addition, the game can be used as a way to attract the attention of people (both children and adults). For example, it can be used as a presentation of a product or idea, to show how a project is expected to appear, etc.

Escape the garden was created primarily for learning purposes. The goal of the author was to create a VR game, which is easy accessible for almost anyone. This type of games was choosed, because of the big interest in Virtual Reality. Many VR games are very complicated, so the author wanted to make a game, that is intuitive and easy assimilable for anyone.

Creating games is a good way to get more into computer programming and computer graphics. It is not only very beneficial, but it is also very entertaining.

The current paper presents the components of the game, different aspects of virtual reality and information about the engine that it's used.

2 How the game works

2.1 Main functionality

The game is designed to be used with VR glasses and VR controller. A smartphone is placed inside the glasses and the player looks directly at the display from a very short distance, so that is why the lenses that are in the glasses are needed. The controller is needed, because when the phone is inside the VR glasses and the screen cannot be touched. Only two of the buttons of the controller are needed for playing the game. One of them is used for moving the camera forward. The other one is used for interacting with objects, which can be done when the user is on the necessary distance. Interaction includes picking up objects, opening doors, pressing buttons, etc. The direction of the camera is changed when the player rotates his head. Third button is used for pausing the game and exit.

Fig. 1 shows how the game looks like on the display of the phone. The display is divided into two parts, one for each eye. The lenses of VR glasses transform the image so it looks more realistic and they solve the problem of the short distance between the eyes and the display.

In Escape the garden, the player is placed in a garden, surrounded from all sides with walls. The only way out is a door that requires a key. The player faces the challenge of revealing all the mysteries of the garden to find this key. He has to find secret codes and to perform multiple tasks to reach the ultimate goal - to get out of the garden.
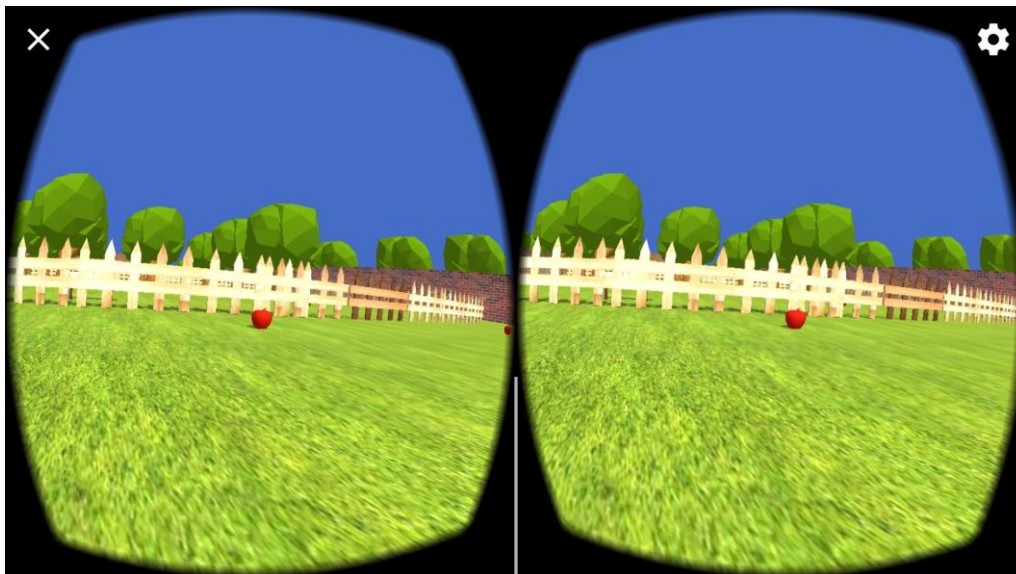

Fig. 1: Game view

2.1.1 Objects and interactions with objects

During his game, the player can interact with the following objects:

- Door - it leads to new challenges. Sometimes a key, secret code, or task execution is needed to open it;

- Signboard – it tells the player what to do at this stage of the game (Fig. 2);

- Door with cipher - to open this door, the player needs to enter the correct code, which is different for different doors (Fig. 3);
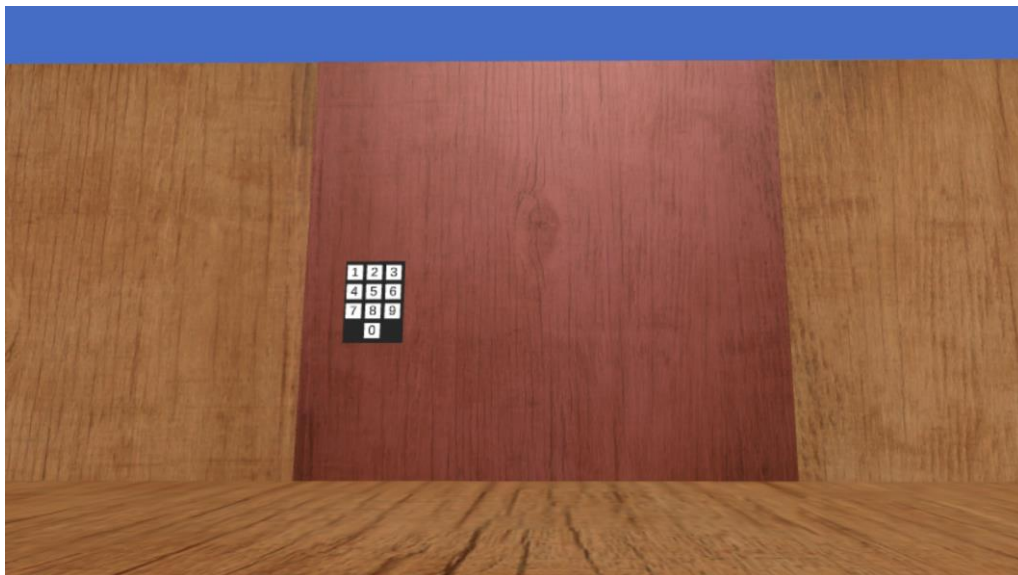

**Fig. 2: Signboard**

Fig. 3: Door with a cipher

- Door with password - to open this door, the player needs to enter a correct word, which may be answer to a certain question or something else (Fig. 4);

- Collectable objects - the player can collect some objects that are needed for a particular purpose;

- Monitor - a display showing a movie that contains some mystery (Fig. 5);

- Remote control - needed to power the TV (Fig. 5);

- Cabinet – it can contain some stuff, to be opened, sometimes it requires a key (Fig. 5).
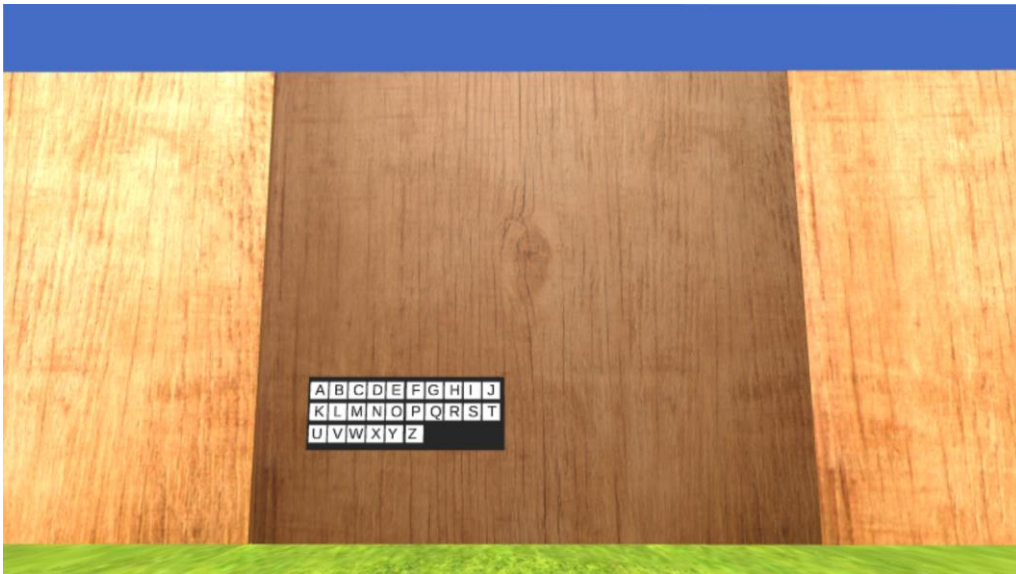
Fig. 4: Door with a password

Fig. 5: Monitor, remote control and cabinet

2.1.2 Main menu and Pause menu

By starting the game, the player will be located at the main menu (Fig. 6) where he can choose to start the game, to go to options or to quit the game. If he chooses to play, he starts from the starting point at the garden. When the game is running, by pressing Esc key the player will be led to the Pause menu where the game is in pause mode. Furthermore, he can choose to resume the game, go on the main menu or to quit the game.

If the player select the options menu, he will be able to change the sound and the brightness levels of the game.

2.1.3 Animations

A lot of animations are used in the game. This is the way of making the objects to move in a particular way. For example, a door is opened by turning it in a certain way. Unity has a built-in feature to make these animations.

Fig. 6: Menu

2.1.4 Architecture of the game

The game is made in such a way, that many of it's objects are customizable. In other words they can be easily used multiple times with different values. For example – each door in the game has it's own variable for the code, needed to open it. This makes the creation of new levels very easy, by reusing the same objects, but with different values.

2.2 Unity engine

Unity engine can be used to make both 2D and 3D games. In order to make a 3D game it is necessary to perform several basic activities. On the one hand, we must create one or several scenes.
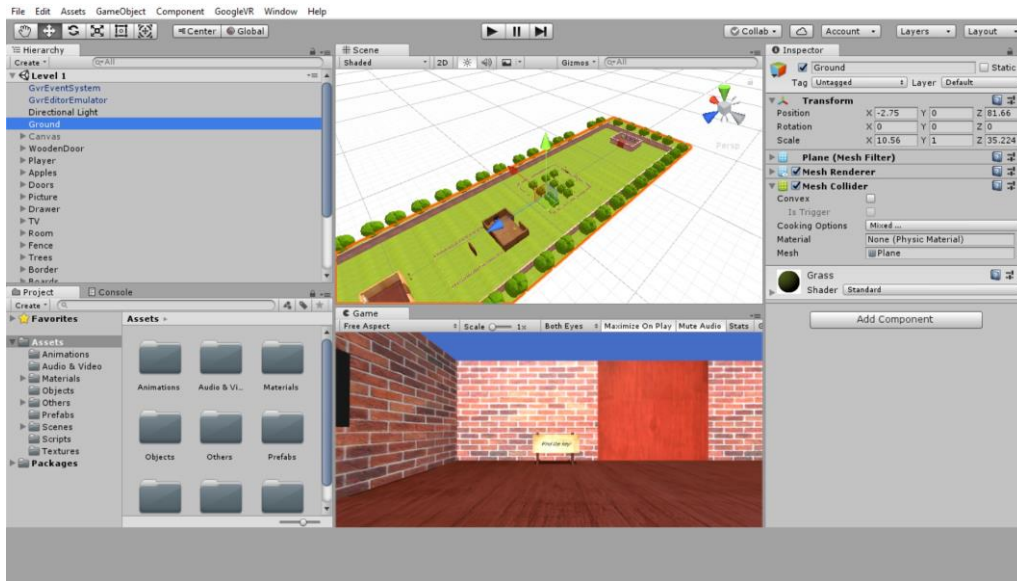
Fig. 7: Unity's framework

At the beginning, each scene is almost empty. The only thing in it is Main Camera. That is the view itself that shows us a part of the scene, when the game is played. Then we add the objects that we want. They can be objects provided by Unity, such as cubes, spheres, cylinders and planes or more complex external objects like people, cars, trees, etc. The external objects can be made on various programs for modeling. One of the most famous programs, which can export 3D models to Unity, is Blender [7].

On the other hand, different characteristics and functionalities are given on these objects. Unity offers us quite a few. For everything that is not provided, we can write C# script, which is bounded to particular object. At the end of the game development, we have multiple objects with multiple scripts attached to them. Unity gives us the ability to multiply our objects and reuse them as much as we want. This way we can easily expand the scene or create new scenes.

The Unity's environment is shown in Fig. 7.

3   Conclusion

Virtual reality can be used for many things in almost any area. The virtual reality is technology that is just emerging. That is why, there are still a lot of challenges for the science in order to expand and improve this technology. There are many tools for achieving this goal and Unity is one of the best of them. It gives us everything we need, in order to create and process 3D scenes.

A lot of challenges have been faced during the process of creating the game. It was hard to find articles on the topic, because it's relatively new. One challenge was to make a user interface for 360 degree virtual world. For this purpose, were used 3D objects instead of UI elements. The biggest problem was to make the game work well on the VR glasses. This was solved by testing it many times.

The author of Escape the garden learned a lot during the process of creation. He plans to expand and improve the capabilities of the game by:

adding new levels, new functionality, new objects, an option for multiplayer and more.

References

[1] Virtual Reality. [online]. available at https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html [accessed 25 March 2019]

[2] Unity. [online]. available at https://unity.com/ [accessed 25 March 2019]

[3] Google VR SDK. [online]. available at https://docs.unity3d.com/2017.2/Documentation/Manual/VRDevices-GoogleVR.html [accessed 25 March 2019]

[4] Google Cardboard. [online]. available at https://en.wikipedia.org/wiki/Google_Cardboard [accessed 25 March 2019]

[5] Oculus. [online]. available at https://www.oculus.com/ [accessed 25 March 2019]

[6] Vive. [online]. available at https://www.vive.com [accessed 25 March 2019]

[7] Blender. [online]. available at https://en.wikipedia.org/wiki/Blender_(software) [accessed 25 March 2019]

Iskren IVANOV
University of Ruse
Department of Informatics and Information Technologies
8, Studentska str., Ruse 7017
BULGARIA
E-mail: iskren_damqnov@abv.bg

Modeling of economic trade processes of enterprise by means of MATLAB simulation tool

Asaf Ali Khan, Mariia Khan

Abstract

In this work economic trade processes of an enterprise were studyied and analyzed.The main trade processes were described in details: purchasing goods, emloyees' payments, ERU, VAT, taxes. The method of the graphoanalytical representation of a mathematical model was successfully applied to modeling of the economic trade processes of an enterprise. The modeling was made with the help of MATLAB simulation tool.

1   Introduction

Recently, economic modeling methods have been increasingly used in economics, which allow to conduct research and analysis of the processes occurring in economic systems, to determine the dependencies between the elements of the systems, to form a scientifically based forecast [1, 2].

Modeling in economics is the reproduction of economic objects and processes in limited, small, experimental forms, in artificially created conditions. In economics, mathematical modeling is often used by describing economic processes by mathematical dependencies. The purpose of mathematical modeling of economic systems is to use the methods of mathematics for the most effective solution of problems arising in the field of economics, using modern computer technology.

Studies of recent years show that the dynamics of economic systems of various levels (industry, production and trading enterprises and other economic objects) are adequately described by multi-tier operator links. Widely used in the theory of automatic control, the method of the **graphoanalytical representation of a mathematical model** of an object in the form of "dynamic links" allows the analysis of a system described by both linear and nonlinear relationships, and can be successfully applied to modeling of economic systems [3, 4].

To determine the adequacy of the method a simulation modeling of trade enterprise MATLAB tool was chosen. The modeling was done with usage of a package of application programs to solve technical problems of computations.

2   Simulation modeling process

The implementation of the model in discrete time required the introduction of the term "memory", which allows you to remember the value of the parameter at its input on the command from the

outside. In the absence of a control signal link retains its output previously memorized state (Fig. 1). On Fig.1 is a scheme in the form of a subsystem "Memory".
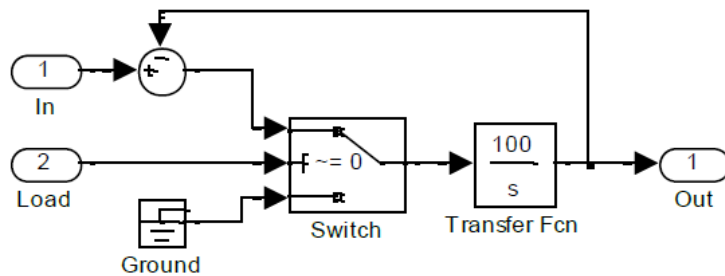

Fig. 1 – The absence of the control signal

The upper state of the "Switch" (in Fig. 1) corresponds to the memory mode of the input signal "In". At the lower state of the "Switch" key, the zero signal is given to the input of the "Transfer Fcn" link, and its output "Out" remains in the state corresponding to the previously memorized value. Control is accomplished by applying a logical signal to the "Load" input. The state of logical "0" corresponds to the data storage mode, the logical "1" corresponds to the storage mode of the input signal.

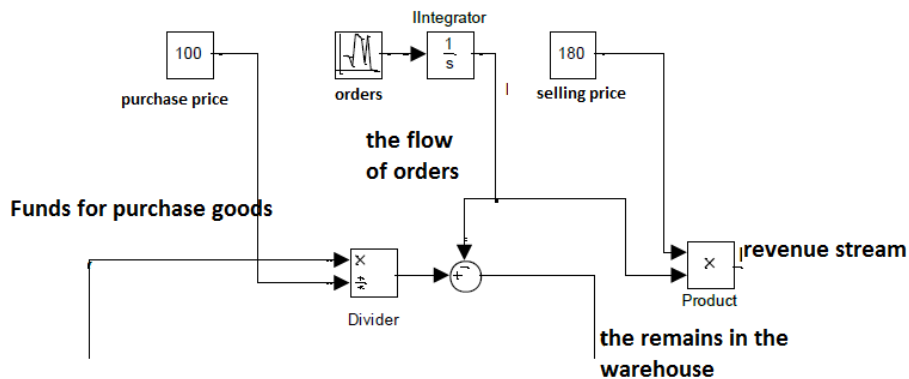Let's start modeling with the function of purchasing goods (Fig. 2).

Fig. 2 – The Modeling the functions of procurement of goods

At the entrance of the scheme comes the flow of funds for the purchase of goods. The "Divider" block calculates the quantity of goods purchased. The flow of orders is randomly generated. Multiplying the flow of orders by the selling price, we obtain the value of the revenue stream, and subtracting the flow of orders from the volume of purchases, we obtain the current value of the remaining goods in the warehouse. From the stream of the proceeds we will deduct funds to pay taxes and pay for employees.

We will form a flow of funds directed to pay for employees. In accordance with the labor law, wages must be paid at least 2 times a month, so we will form two flows of funds corresponding to the advance (20th of the month) and wages (5th of the month). [3]

Since the scheme consists of two identical branches, consider its work on the example of the upper branch. The block "Counter Limited" performs the function of the counter of days of the month. At the output of the "Compare to Constant" block, a logical "1" signal is generated for a duration of 1 sec. (1 working day) every time when the reading of the counter of "Counter Limited" block coincides with the constant of the comparison unit. At the same time, a switch equal to the sum of the employees' monthly salary is sent to the Integrator input "Integrator1" for 1 second via the "Switch" key. The signal at the integrator output "Integrator1" in this case becomes equal to:

$$Y(t) = y(t) + \int_0^1 3000\, dt = y(t) + 3000$$

Summing up the flow of funds allocated for salary and prepayment, we will get the total flow of funds for employees' salaries (Fig. 3).
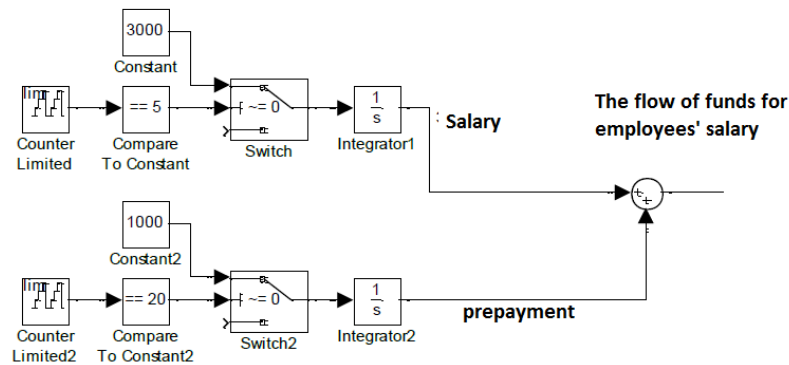

Fig. 3 – The total cash flow

Next we will calculate the amount of unified social contributions paid from the wages of employees. Payment of the single social contribution (ERU) manufactured in the period up to the 15th of the month following the settlement. [4] Therefore, we can fix with the help of the previously created subsystem "Memory" the value of the wage flow as of the 15th day of the month. Multiplying the output signal by the total tax rate (on average about 30%) we will get the amount of the flow of funds allocated for the payment of ERUs (Fig. 4).
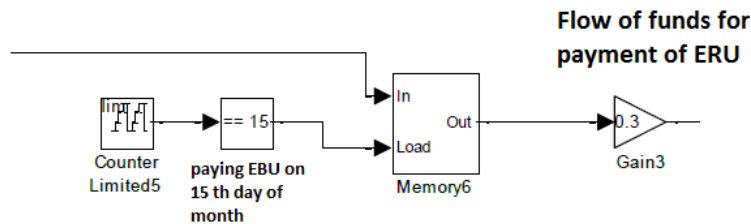
Fig. 4 – The flow of funds for payment of ERU

We will simulate the mechanism of calculation and payment of value added tax. VAT is paid at the end of each tax period based on the actual sale (transfer) of goods for the expired tax period in equal installments not later than the 20th day of each of the three months following the expired tax period. VAT deductions are subject to the amount of tax presented to the taxpayer when purchasing goods (works, services) in respect of goods purchased for resale. The model implementing the described mechanism of tax calculation is shown in Figure 5.

Fig. 5 – The model of calculation and payment of VAT

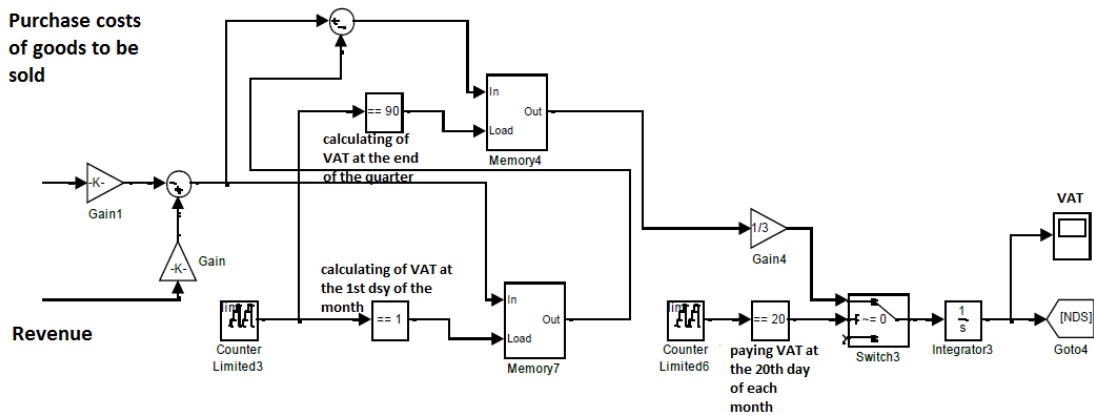The "Gain" and "Gain1" blocks separate the amount of VAT from the amount of revenue and the cost of purchasing goods at the rate of 18/118. A scheme based on the "Memory7" and "Memory4" memory blocks and the "Counter Limited6" calendar (with a period of 90 days) forms the output value accrued over the VAT quarter. Integrator "Integrator3" together with the blocks "Switch3", "Gain4" and calendar, based on the block "Counter Limited6" (with a period of 30 days) generates values equal to the sums subject to monthly payment in equal parts during the tax period. Let's make a model of accrual and payment of income tax (Fig. 6).
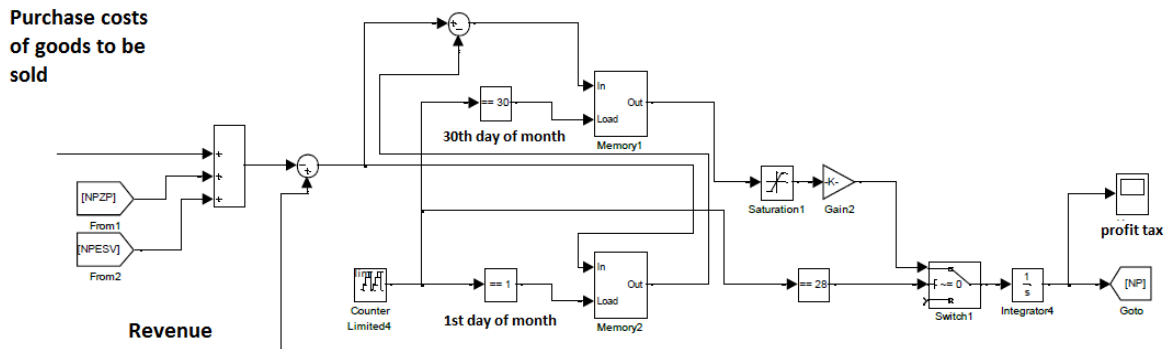

Fig. 6 – Model of calculation and payment of income tax

Input adders implement the mechanism for the formation of the tax base for income tax. Profit tax is deductible for purchases of goods purchased for resale, labor costs, expenses for the payment of ERUs. A scheme based on the memory blocks "Memory1" and "Memory2" forms at its output a value equal to the sum of the tax base calculated for the period from 1 to 30 of the month. On the basis of the "Memory" element, a node has been assembled that implements the function of paying income tax on the 28th of the month following the settlement one. The coefficient "Gain2" is equal to the tax rate of 20%.

The model of goods procurement management is shown in Figure 7. The model describes the management of funds allocated weekly for the purchase of goods, taking into account the limited capacity of the warehouse.
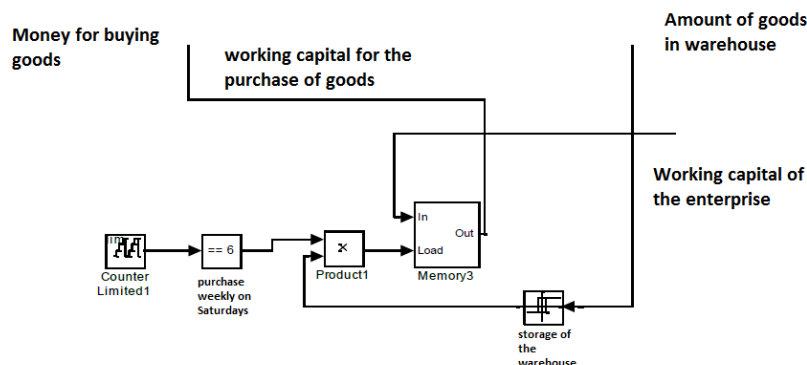

Fig. 7 – The management model of procurement of goods

3 The results of the modeling

The simulation results are shown in Fig. 8. The initial accepted "zero" state of the system – no investment, working capital for the purchase of goods come only from buyers (advance payment for the goods).
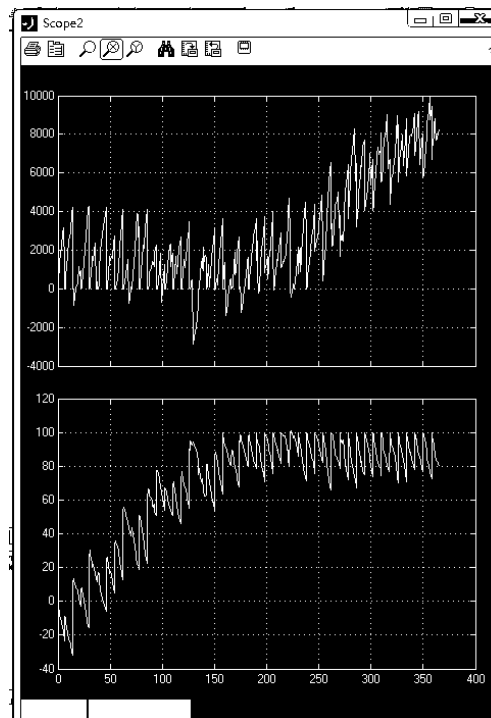
Fig. 8 – Results of the modeling

The upper graph represents the amount of working capital in the company's account after deducting mandatory payments. The lower graph shows the number of goods in stock minus the already sold (prepaid) goods.

Analyzing the graphs presented in Fig. 8, we can make the following conclusions:

a) in the initial period of enterprise operation there is a high risk of "cash gaps" insufficiency of working capital allocated to the repayment of current debt (upper graph);

b) at the beginning of the work of the enterprise with the specified parameters of the model, the amount of orders may exceed the enterprise's ability to purchase goods (processing the order queue), so there will inevitably be a delay in the delivery of goods to buyers;

c) at a given average density of the order flow during the first 200 days, the funds will be directed to increase sales turnover, and then, when the number of goods reaches the warehouse capacity threshold, the enterprise will begin to receive net profit.

Figure 9 shows the cumulative graphs (in order from top to bottom) of wages, deductions for ERUs, VAT, income tax.

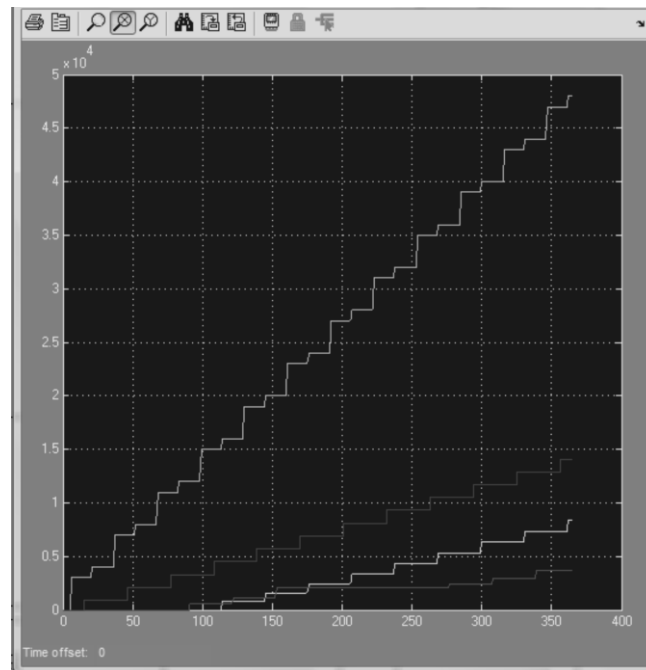Fig. 9 – Graphs of wages and taxes

From the presented graphs it is seen that the main share of costs in the developed model is funds allocated for labor. Drawing attention to the profit tax schedule (lower graph in Fig. 9), we can see that the company is on the verge of profitability.

4   Conclusions

The developed model, despite the rather large number of accepted simplifications, allows simulating economic processes taking into account daily dynamics. The model is scalable and can be supplemented with modules that take into account other parameters of economic processes, not only trade enterprises, but also production, consulting, etc.

References

[1] Tsarkov V. A. *Dynamic models of economic systems*, Moscow: Audit and financial analysis, 2005.
[2] Tsarkov V. A. *Modeling of economic dynamics of the enterprise*, Moscow: Audit and financial analysis, 2011.
[3] Henry, J. and Walker, D. *Managing Innovation*, London: Sage Publications, 1991.
[4] Basalla, G. *The Evolution of Technology*, Cambridge: Cambridge University Press, 1988.

Asaf Ali KHAN
Siena State University
Department of Business and law
Piazza San Francesco, 8, Siena
ITALY
E-mail: asafalikhan8@gmail.com

Mariia KHAN
Siena State University
Department of Information Engineering and Mathematics
Via Roma, 56, Siena
ITALY
E-mail: mariiakhan@yandex.ru

Applying Gamification to an Information Security Awareness E-Learning Platform

Dominik G. Klopfer

Abstract

This paper describes the design and the development of an E-Learning platform with gamification approaches for security awareness training. The platform can be used to provide content within a serious context, which gets communicated to the user with playful game design elements. The methods used are illustrated by examples and explanations. The special feature of this E-Learning platform is the simple and intuitive design, as well as the usability on different devices through the responsive presentation of the platform.

1 Introduction and Motivation

The german domestic intelligence service "Bundesamt für Verfassungsschutz (BfV)" is currently warning about targeted cyber-attacks on universities and research institutions [1]. With these attacks, which probably have an intelligence service background, the attackers are trying to gain access to university networks and confidential university data. The BfV assumes that employees and students are specifically selected as victims. Strict rules to increase the security of information technology are difficult in an academic environment, because strict rules and guidelines do not achieve the desired results, especially for students. Various experts come to this conclusion in a report of the newspaper "The Guardian" [2]. Against this background, it is important to train the security awareness of the target group with playful elements in order to bring the urgency of the topic closer to the students.

For this reason, we have designed a university E-Learning platform with gamification elements. A special focus of the platform is on security awareness training. The platform should not have a classical conservative academical character but should awaken the play instinct and convey valuable information to the learners in a playful way. The gamification concept is based on the insight that people learn and work more easily when they have fun. Therefore, gamification integrates playful elements into productive or other processes in order to make the results of the process appear less forced. Instant gratification keeps learners motivated.

2 University Project

Within the module Software Development Project, students can choose a topic of interest and implement a software solution to this topic within one semester. Our project team consisted of five computer science students and one business information systems student. Our project team

decided to design and implement an E-Learning platform for information security awareness that follows gamification approaches.

The requirement of Prof. Dr. *Kristin Weber*, our lecturer in charge, was a modern design of the platform in contrast to the existing university E-Learning platform. The platform should not have the classical and conservative character of a university but should awaken the play instinct and convey valuable information to learners in a playful way. The project team used agile software development methodologies and consciously distinguished itself from a conservative project organization, such as the waterfall model. The Scrum framework [3] was used for this implementation, which was realised with Jira, an agile project management solution from Atlassian.

3   Background

3.1 Information Security Awareness

Information security awareness is a research area within the broader information security context. The users of information technologies should apply security awareness in practice.
Hänsch & Benenson [4] present three views on security awareness:

- Perception: students recognise a threat.
- Protection: students know how to deal with the threat.
- Behaviour: students know what a threat is, what they can do about it and behave accordingly.

Increasing awareness means that the user become aware of the impact and immense damage caused by information technology risks. Information security compliant behaviour is the result of a person's knowledge, habit and intention to behave in relation to a particular behaviour [5]. An information security compliant behaviour can be, e.g., the use of different and secure passwords.

The most complex and influential factor can be described as the intention to behave. This is made up of various emotions and beliefs. Despite a positive security awareness, however, the final execution of a behaviour can be prevented by organizational restrictions in the environment of a person. With the knowledge of the individual factors of a person it is possible to influence them in order to change an undesirable behaviour or to strengthen a desired behaviour. There are several possible interpretations in the field of security awareness. Some views refer only to the dangers of the Internet. A large area is the link to privacy and IT security issues. Awareness is generally associated with risk management and information security. Those responsible must define how the issues are understood and which categories they are assigned to. It is not possible to look at the topic in general. It must always be placed in the context of the respective institution.

3.2 E-Learning

E-Learning is based on Web Based Training (WBT) [6], that is, interactive learning programs that are processed in multimedia over the Internet. Web Based Trainings allow for retrieving knowledge and for consolidating it through exercises and transfer tasks. In a self-organised learning process, WBTs can be used to support the structuring of individual learning processes. This enables learners to provide the necessary orientation and learning opportunities. In addition, information is provided to the users, which gives insights into their learning process and progress and thus promoting activating learning strategies. These aspects have proven to be essential success factors for learning.

3.3 Gamification

Gamification is the playful design of activities that are used with game design elements in a non-gaming context [7]. A non-gaming context means that they are not designed just to entertain

players, but to reach other goals. For example, the use of gamification in the area of information security is used to train security awareness. Furthermore, Gamification is intended to trigger a playful experience for the user. Game elements are all objects users come directly in touch with, such as points, badges, rankings or objects. These elements are explained in detail below. They fulfil both the feedback function and a performance motivation. The feedback function is designed to give positive or negative feedback on the answer to a question. Performance motivation, for example, is triggered by the desire to receive a badge. As listed above, there are several ways to use elements of gamification. Diversification and a combination of different approaches not only prevent monotony, they also offer the user a varied gaming experience. When implementing gamification, we took inspiration from Hsin-Yuan Huang & Soma's 5-step process [8] for applying gamification in education (cf. fig. 1).

Fig. 1: 5-steps process gamification (based on Huang and Soman, 2013)

The first step is to determine the target group. In our case, these are students. It helps to focus the platform on the target group. Because we are a faculty of computer science and business informatics, we do not have to learn from scratch because there is already a basic knowledge about security awareness. The second step is to define the desired achievement of the learning program. We want to increase the security awareness of the students on our platform. The third step is how to structure the experience. In our case, we use levels to quantify the knowledge. This gives the students a good orientation on how much experience points (EXP) they need to reach a higher level. This way they always have the goal in mind. In the step of identifying resources, you think about which rules are implemented on the platform. We have decided to implement test sheets. In addition, you determine what the feedback should look like. With our platform, you receive the feedback after completion of the test sheet. The feedback will explain the incorrectly answered questions. Finally, the five-step process considers which elements should be used for gamification. On our platform, we have tried to increase the pleasure by using different elements. This is in more detail explained in the chapter Realization of Gamification.

4   Realization of Gamification

The following section describes why and how we applied the gamification elements in detail. On our E-Learning platform a student answers questions about the topic security awareness. The topic is divided into several categories. Within the categories there are several questionnaires (cf. fig. 2).

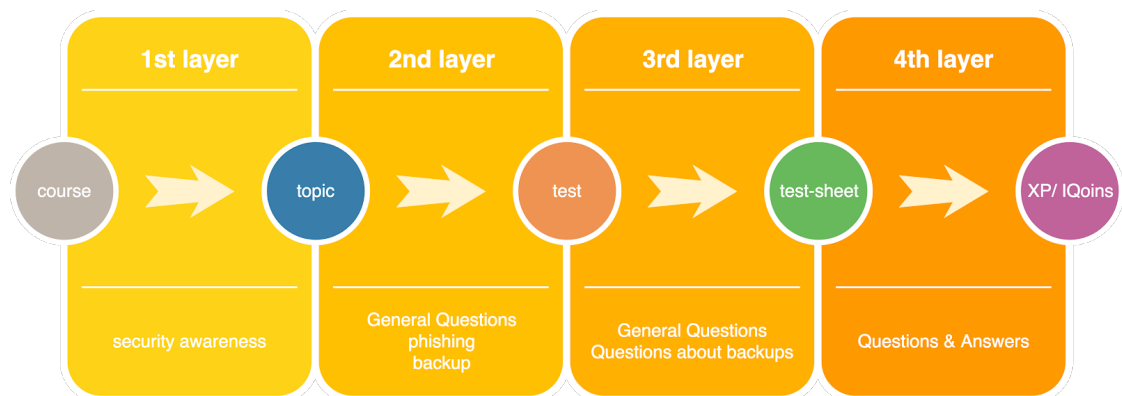| 1st layer	2nd layer	3rd layer	4th layer
course → topic	test	test-sheet	XP/ IQoins
security awareness	General Questions phishing backup	General Questions Questions about backups	Questions & Answers

Fig. 2: Design of the test area on the platform

In order to encourage students to learn, we have devised an experience system/level system. After the successful completion of a test-sheet, the users receive EXP. If the users have accumulated enough EXP they reach a new level on the platform. How many EXP the users receive depends on the difficulty of the test and the test mode.

The so-called "game mechanics" are divided into two areas, personal elements and social elements. Personal elements are elements that satisfy the intrinsic desire for more, such as levels, achievements, currencies. Our social element is the ranking. In this area, the user compares his/her progress and success with that of others. This can spark a spirit of competition in the community. Gamification has a direct impact on commitment and motivation and indirectly leads to more knowledge and improved skills [7]. We have used various methods to increase the competition will and motivation. In the following, we describe which game mechanics we have installed and what our intention was to use them.

4.1 Ranking

On an EXP ranking list, the users can compare themselves with others. This promotes competition and ensures that the users "push" each other to a better performance. Ranking allows for comparing one's own success with that of others. This comparability should increase the competitive spirit of the players and motivate them to new top performances. Rankings can be constructive forms of trigger competition, set clear goals and give feedback [8].

FHWS Learning Plattform — Courses | Ranking | Webshop | Tom ▼

#2 Albert 6500 XP
#1 Lisa 9100 XP
#3 Reiner 5555 XP

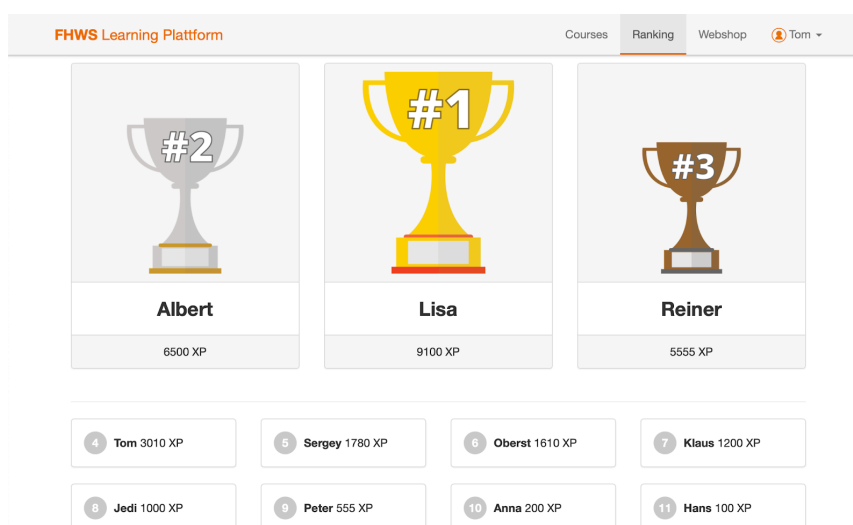| 4 Tom 3010 XP | 5 Sergey 1780 XP | 6 Oberst 1610 XP | 7 Klaus 1200 XP |
| 8 Jedi 1000 XP | 9 Peter 555 XP | 10 Anna 200 XP | 11 Hans 100 XP |

Fig. 3: Representation of the ranking

4.2 Level system

All users start their journey on level 1. The more tests they complete and the better they answer these, the more EXP they receive. After reaching a threshold, the level rises. The maximum level on our platform is currently level 10, but it can be increased, if necessary. To prevent "hardcore users" from not being caught by "casual users", the required experience points for a level up are doubled per level. Empirical studies have shown that levels promote satisfaction with the platform and a positive assessment of quality and performance [8]. Levels are another tool to measure the commitment of the users.
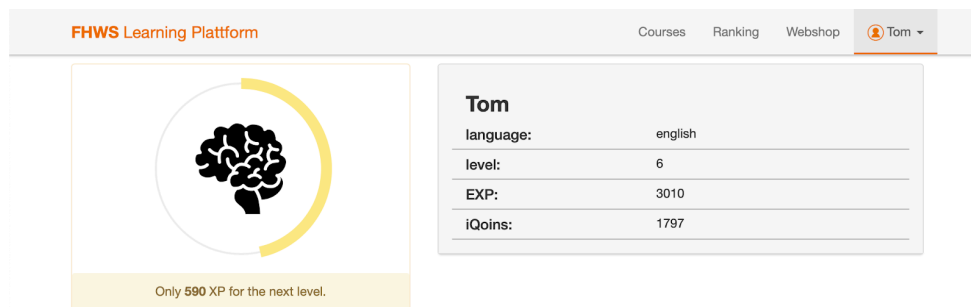
Fig. 4: Level view in the user area

4.3 iQoins (currencies)/ EXP

We have established our virtual currency iQoins (iQs) in order to suggest that the users receive not only a monetary reward for doing tests, but also a mental one in the form of knowledge acquired. As with EXP, the amount of iQs received depends on the test mode and the difficulty of the test. iQs obtained can be spent on items in the shop. EXP are a basic component of all gamification applications. One of the most important functions of EXP and iQoins are their feedback function [8].

4.4 Achievements

Achievements are another form of reward for the user. These are rewarded for special or unusual achievements. Achievements are displayed via badges in the user's profile. With achievements we want to create a competitive character so that the users broaden their horizons while looking for new achievements. Achievements can be used to control goals as well as the general behaviour of the user [8]. These achievements also give positive feedback to the user.
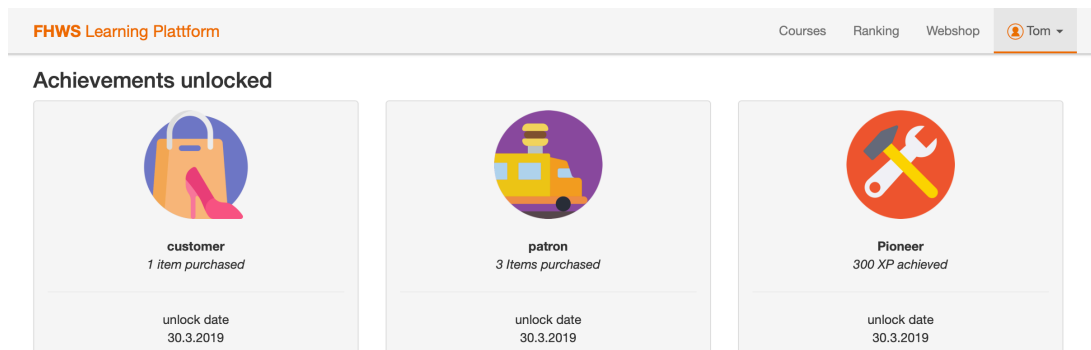
Fig. 5: Unlocked achievements in the user area

4.5 Webshop and items

In the webshop we offer various items that the users can purchase for iQs. At the beginning we offer "Booster" that multiply, if purchased, the EXP or iQs received during the tests. So far, there is no scientific basis for this. This is based on the assumption that the webshop makes the platform more interesting. At this point, the effects of the webshop would have to be analysed by interviews with test persons.
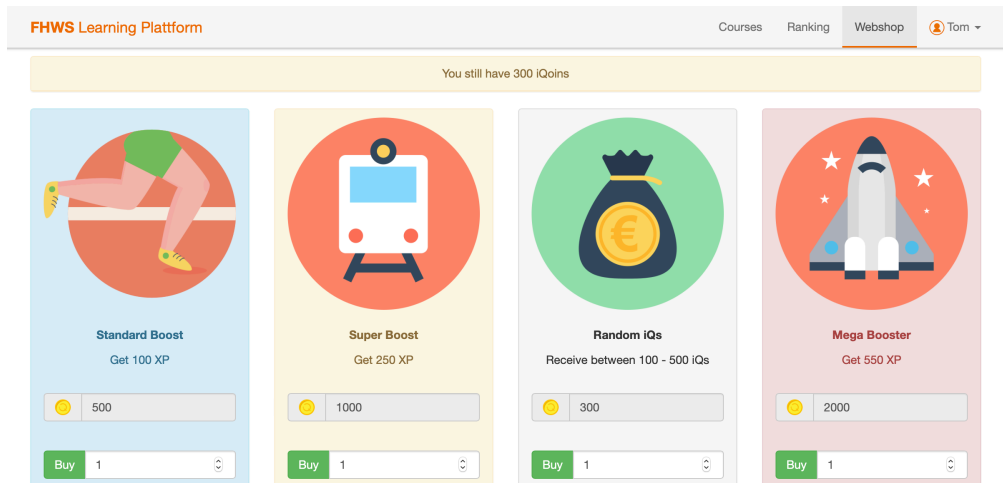

Fig. 6: Illustration of the platform webshop

5 Platform Architecture

5.1 Frontend

5.1.1 Bootstrap

The frontend is implemented in HTML, CSS, and JavaScript. The frontend framework Bootstrap [9] was used to simplify the process. The framework offers a library of different elements, for example button groups, checkboxes, navigation bars or panels. Another advantage of frameworks is that with their usage a responsive web design is often supported in the frontend without additional development work. There are a number of free frameworks available, such as Foundation, Gumby, Bootstrap, or Helium. Due to the great popularity of Bootstrap, we decided to use this framework. To use Bootstrap, you either need to download the library and embed it into the development project or embed an HTML link to use the framework.

5.1.2 JavaServer Pages (JSP) Files

We used JavaServer Pages (JSP) to embed Java code into HTML for the project. JSP is a web programming language for generating HTML output from a web server. The following example is the navigation bar *nav.jsp*.

```
<ul class="nav navbar-nav navbar-right">
   <% if (user != null) { %>
   <li class="${pageContext.request.requestURI eq '/saplattform/course.jsp' ? ' active' : ''}">
      <a href="course.jsp">Course</a></li>
   <li class="${pageContext.request.requestURI eq '/saplattform/ranking.jsp' ? ' active' : ''}">
      <a href="ranking.jsp">Ranking</a></li>
   <%//} %>
</ul>
```

5.1.3 JavaScript / jQuery

We used JQuery to evaluate the user interaction and to change/reload/generate content. This enabled us to use dynamic HTML. The following line adds a text to an HTML tag with the class "exercises". The following line is inserted in *admin.js*.

```
$('.exercises').append("<p>Exercise1</p>");
```

5.1.4 Ajax Request

To create an asynchronous communication with webservice classes, we used Ajax Request. This is a concept of asynchronous data transfer between a browser and the server. This allows us to create a request and process incoming responses. For example, insert the incoming JSON object into an HTML tag. As an example, you can see the use of the Ajax Request in *generalFunction.js*.

```
function ajaxRequestHandler(type, url, data, successFunc) {
  $.ajax({
    type: type,
    url: url,
    data: JSON.stringify(data),
    dataType: 'json',
    contentType: "application/json",
    success: function (jsonObject) {
      console.log(jsonObject);
      successFunc(jsonObject);
    },
    error: function (jqXHR, textStatus) {
      console.log(jqXHR);
      console.log(textStatus);
    }
  });
}
```

5.2 Backend

5.2.1 Web Service

We use a web service that communicates with a client (e.g. user with a web browser) and receives or sends data. This transfer of data is also known as a request/response. The web service runs on a web server. A Web service uses Web service classes and their methods to react to the respective request type. For example, a Web service class "Item" has a method that gives the client all items. So data can be sent once by parameter, e.g. items/{itemID}, then the placeholder itemID would have a number. Or the client can also send something in the body. These data are then JSON objects and can be converted into Java objects with Genson (see 5.2.2 Genson) to use them for further processing. When everything is done, the method can send a Java object to the client so that it can be displayed in HTML. At this return, Genson converts this Java object into a JSON object and sends it to the client.

5.2.2 Genson

To enable a converting from a JSON object to a Java object, we used the library Genson. For this, a JSON object is mapped one to one into a Java object, i.e. all fields in this JSON object must have the same name as the variables in the Java object. Furthermore, the Java class must have a default constructor and the setter and getter methods, because Genson first initializes an empty object and fills it with the setter methods.

5.2.3 Database

For the project creation, we used the bottom-up principle. This means that smaller sub-problems are solved first and then, with their help, larger problems are solved until the overall problem is

solved. That's why we have dealt with the individual database tables and the ERM, because we could get a detailed view of all objects (e.g. user, courses, test). Building on this, we were able to create our classes and later our web services.

5.3 Communication

Communication on the client side is via JavaScript (e.g. button press and data transfer) and Java runs on the server side. Therefore, we need a transfer of the data between client and server. For this transfer we use JSON. With JSON, the data can be nested and sent via AJAX via a specific URL to the server, which receives and processes this data via web service classes. This generates a response, which is then intercepted by Ajax and inserted into the HTML content using jQuery.
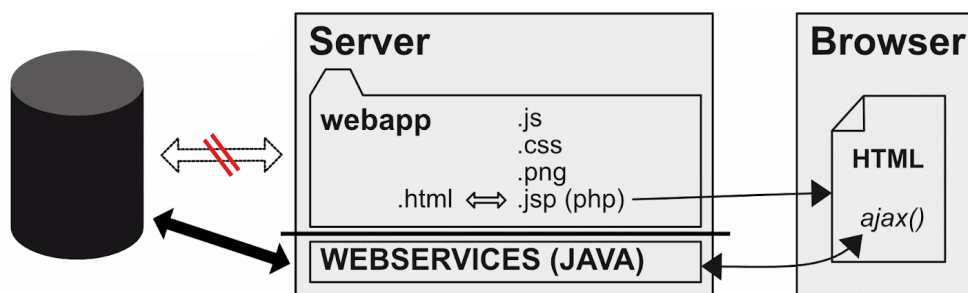
Fig. 7: Communication between server and browser

6 Application

6.1 Administration View

The complete control of the E-Learning platform is done in an administration frontend. Here an administrator can set achievements, create items in the webshop, read user statistics, as well as create and edit questions. The administrator can also upload resources in this area e. g. icons for achievements or courses. In the user administration area, different permissions can be set, for example Content Manager, to create tests (cf. fig. 8). The various authorizations can be set in the user management area.

Fig. 8: Usermanagement view of an admin

In the following we will show how a user who has the rights as content manager can create a test and how the test is executed. In the courses section, all courses and topics are listed (cf. fig. 9).
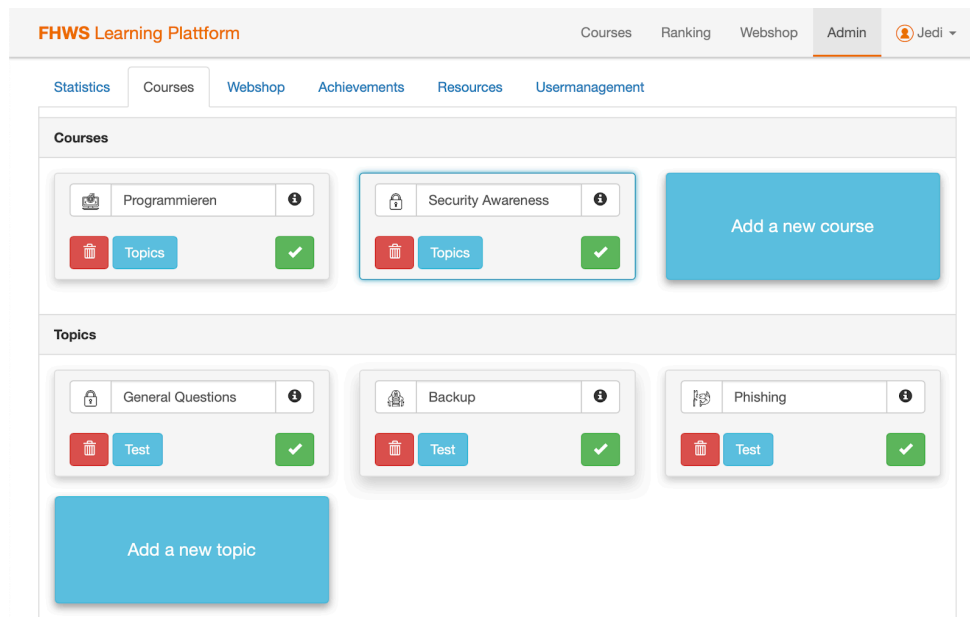
Fig. 9: Admin overview of courses and topics

The admin selects a topic to edit the test. For the time being we have only implemented a Questions & Answers function (cf. Fig. 10). Further, speed tests are planned that enable users to perform the test within a given time.
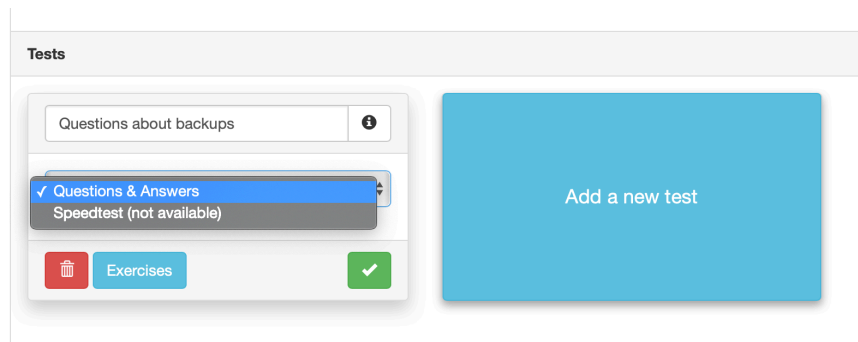
Fig. 10: Setting the game modes

Within the test sheet, the administrator is able to create questions. Questions can be one choice or multiple choice. It is also possible to set the difficulty for the question. When correctly answering more difficult questions the users gets more EXP and more iQoins (cf. Fig. 11).
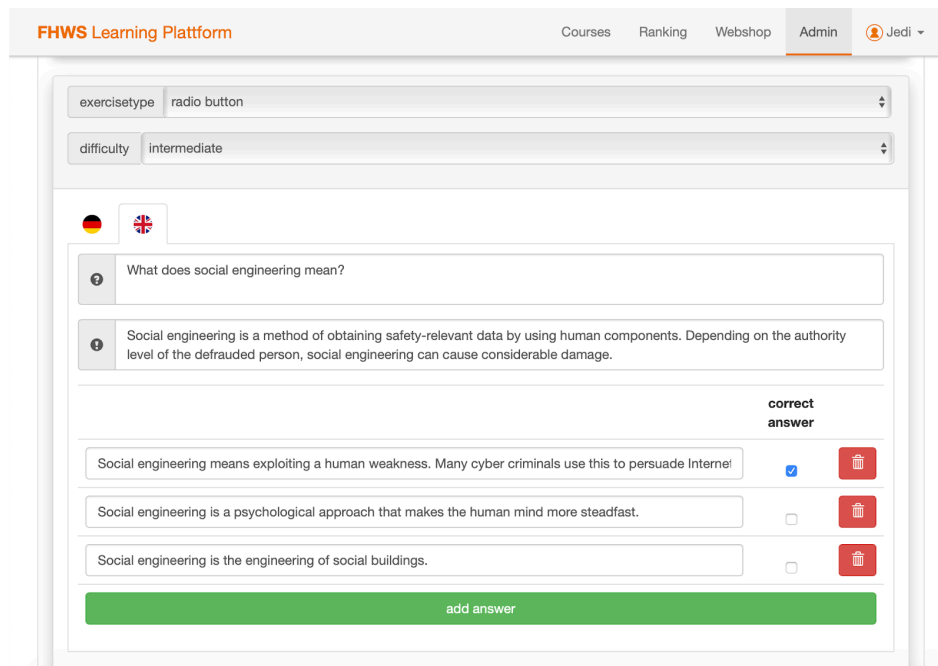
Fig. 11: Question creation view

6.2 User View

The user sees all courses in the courses section. In the test mode, the user can practice before taking a test sheet (cf. Fig. 12).
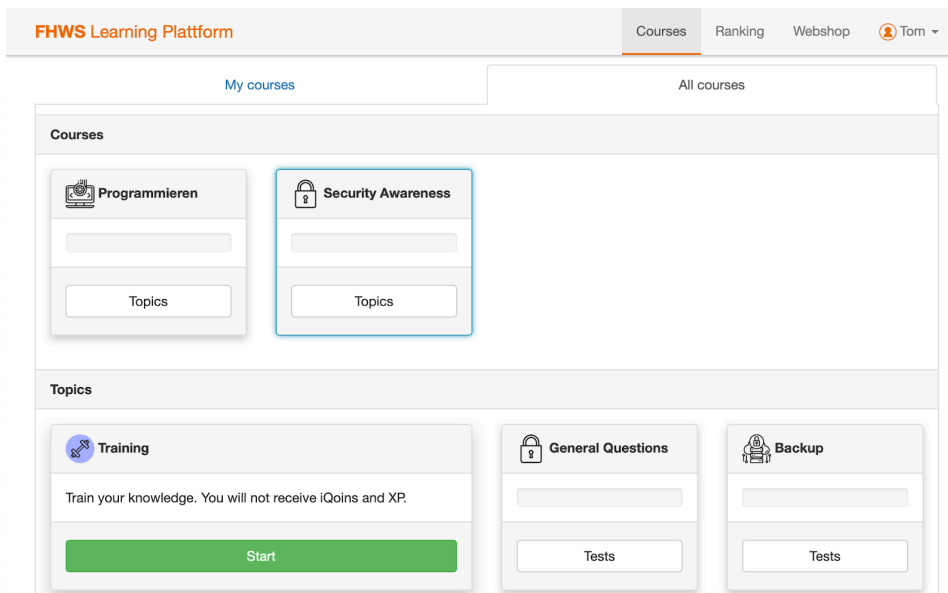
Fig. 12: Overview of courses

In the test the user proceeds with the questions by navigating with the "previous" and "next" button. After all questions are answered the test is submitted (cf. Fig. 13).
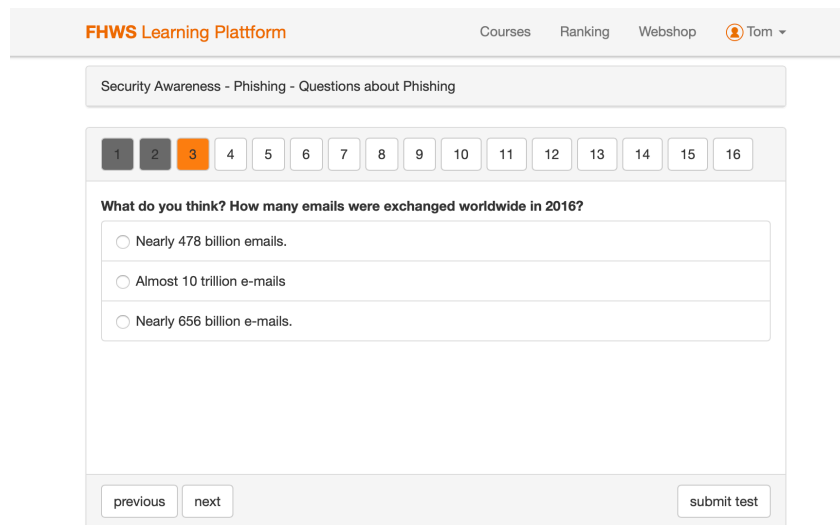
Fig. 13: Insight and navigation in a test

In the figure (cf. Fig. 14) the user did not answer all the questions correctly and did not pass the questionnaire. Nevertheless, they received iQoins and experience points. Users can repeat test sheets as often as they want.
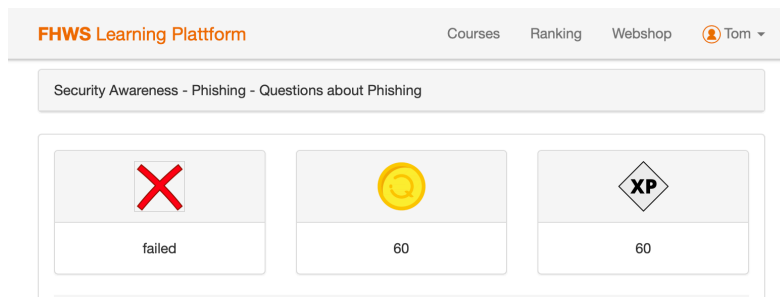
Fig. 14: Result after a test

After completion of the test, right and wrong answers are displayed as well as explanations of the questions (cf. Fig. 15).
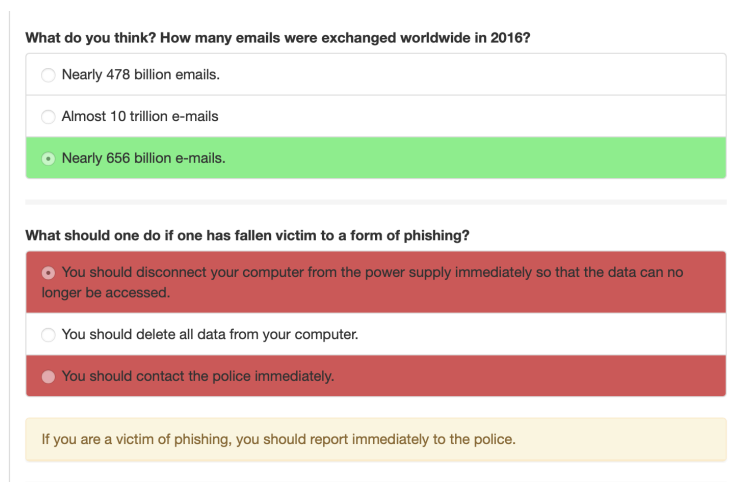
Fig. 15: Feedback of the given answers

7 Conclusion and Future Development

The developed E-Learning platform laid the foundation for an E-Learning platform using gamification aspects. Overall, the software development project was very extensive, as we started on a greenfield site and had to familiarize ourselves with the respective topics. Due to time constraints, we were unfortunately not able to implement everything we had planned, but we are nevertheless very satisfied with our performance and the end result. With the help of Bootstrap, we have created a modern and responsive learning platform that is intuitive and doesn't require long instructions. But the development does not end here. There is still a lot to be done on the project so that more gamification approaches can be pursued. The speed test, for example, where the user only has a limited time to answer the questions. As well as further achievements for the motivation of the users. Unfortunately, the long definition phase of the project didn't allow this any other way. The teamwork worked very well, and everyone contributed wherever they could. The development was great fun for us, it challenged and encouraged us. In our eyes, the project was a complete success. It still has to be tested whether the webshop actually increases users' motivation. In addition, the developed E-Learning platform must be tested over a longer period of time in order to obtain empirical results as to whether it contributes to a change in the security awareness of the user.

References

[1] *"Nachrichtendienstlich Gesteuerte Cyberangriffe."* Bundesamt Für Verfassungsschutz, 6 June 2018, www.verfassungsschutz.de/de/oeffentlichkeitsarbeit/publikationen/pb-cyberabwehr/broschuere-2018-05-nachrichtendienstlich-gesteuerte-cyberangriffe.

[2] Tickle, Louise. "*How Strict Is Too Strict at School?"* The Guardian, Guardian News and Media, 18 Sept. 2017, www.theguardian.com/education/2017/sep/18/how-strict-is-too-strick-great-yarmouth-academy-smile.

[3] Atlassian. *"Scrum - What It Is, How It Works, and Why It's Awesome."* Atlassian, www.atlassian.com/agile/scrum.

[4] Hänsch, N., Benenson, Z.: *Specifying IT security awareness.* In: Morvan, F., Wagner, R., Tjoa, A. (Hrsg.): 25th International Workshop on Database and Expert Systems Applications, Page 326-330. IEEE Computer Society, 2014.

[5] Schütz, A.E.: *Information Security Awareness: Its Time to Change Minds!* In: Proceedings of International Conference on Applied Informatics Imagination, Creativity, Design, Development - ICDD 2018.

[6] Erpenbeck, John, et al. *E-Learning Und Blended Learning: Selbstgesteuerte Lernprozesse Zum Wissensaufbau Und Zur Qualifizierung.* Springer Gabler, 2015.

[7] Wendy Hsin-Yuan Huang and Dilip Soman. *A Practitioner's Guide To Gamification Of Education.* Research Report Series Behavioural Economics in Action, 2013.

[8] Sailer, Michael. *Die Wirkung Von Gamification Auf Motivation Und Leistung.* Springer Fachmedien Wiesbaden, 2016.

[9] Otto, Mark, and Jacob Thornton. *"Bootstrap."* · The Most Popular HTML, CSS, and JS Library in the World., getbootstrap.com/.

Dominik G. KLOPFER
University of Applied Sciences Wuerzburg-Schweinfurt
Faculty of Computer Science
Sanderheinrichsleitenweg 20
97074 Würzburg
GERMANY
E-mail: dominik.klopfer@student.fhws.de

Improving Camera-based Document Analysis with Deep Learning

Nikolai Körber

Abstract

Capturing document images with a smartphone provides a convenient way to digitize physical documents and to facilitate the automation of document processing and information retrieval. In contrast to flatbed scans, camera-captured documents require a more sophisticated preprocessing pipeline, because of perspective distortion, uncontrolled lighting and physical deformation.

The goal of this work is to solve one major challenge in the domain of camera-based document analysis, namely perspective correction, by means of a Deep Neural Network. We investigate different neural network architectures on a large-scale synthetic dataset in order to estimate the document's corner points from a single input image. The distorted image is then mapped to its canonical position by using the 4-point homography parameterization. The best result is achieved by a modified Xception-network, with a mean displacement error of 3.38px. Finally, the correction component is integrated into Tesseract 4.0 and evaluated on the SmartDoc 2015 challenge 2 test set. Our experiments show that the correction component improves the character accuracy results by more than 15 percentage points (93.11%), in comparison to Tesseract by itself (77.27%).

1 Introduction

Automated document processing, regarding e.g. invoices or travel expenses, can substantially increase a company's productivity and reduce its operating costs. Traditionally, Optical Character Recognition (OCR) tools have been developed in order to extract relevant textual information from scans using flatbed scanners. Nowadays, with increasing popularity of smartphones, people tend to digitize physical documents via mobile cameras. This change is mainly due to the ease of portability and reduced costs, since no special hardware device is required. The drawback, however, is the additional effort required to cope with perspective distortion, uncontrolled lighting and even folded or curved documents.

The right hand side of figure 1 illustrates a simplified OCR workflow. It consists of four main phases which build upon each other:
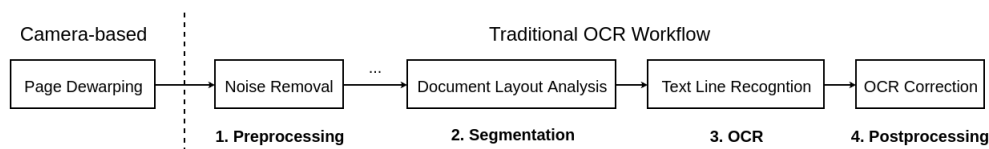

Figure 1: Extended OCR Workflow

First, the input image is preprocessed, including rescaling, binarization, rotation, etc. Next, the document image is divided into logical segments, like e.g. images, graphics, and text. The text regions are further divided into text lines, ideally arranged in the correct reading order. In the following stage,

text line images are converted to machine interpretable strings. Finally, those strings are corrected by means of probabilistic models over sequences of characters (language modeling).

In order to be able to employ traditional OCR tools (e.g. Tesseract OCR [34]) using mobile scanners, the pipeline is extended (see the left hand side of figure 1) by a page dewarping component. The goal of this extension is to transform the camera-captured document such that it is almost indistinguishable from a flatbed scan.

In this work, physical documents are assumed to lie in a plane, e.g. on a table. That is, they do not contain folds or curves, and thus page dewarping reduces to the task of perspective correction.
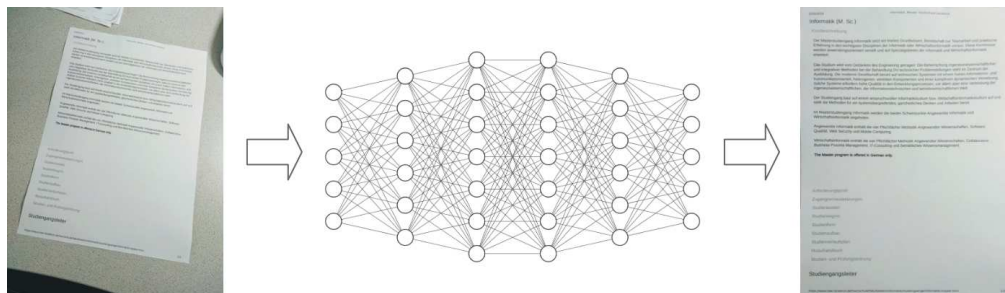

Figure 2: Deep Learning based Page Dewarping

We would like to make the following contributions: i) A perspective correction component is trained on a large-scale synthetic dataset, based on the ideas presented in [9]. In contrast to that work, however, corner points outside of the image boundary are learned in a more realistic fashion. ii) The trained model is plugged into Tesseract 4.0 [7]. It is evaluated on the SmartDoc 2015 challenge 2 test set, by using the official evaluation tools provided by the competition organizers. iii) The source code for both the synthetic dataset generator and the extended OCR pipeline will be published on Github [26].

The rest of this paper is organized as follows: Section 2 reviews recent page dewarping methods for camera-captured documents, while section 3 describes the proposed methods in detail. In section 4, we detail publicly available datasets for benchmarking and our procedure to generate synthetic datasets. Section 5 presents our experiments, and section 6 concludes this paper.

2   Related Work

According to [33], dewarping approaches can be categorized into two groups: i) 3D shape reconstruction of the page by using specialized hardware, e.g. stereo cameras, structured light sources or laser scanners. ii) reconstruction of the page by using a single camera in an uncontrolled environment. Our work focuses on the latter approach, since the requirement of specialized hardware reduces the flexibility of camera-based approaches. In the following section, recent perspective correction methods from a single input image are reviewed.

Traditional approaches extract clues about the document layout, such as document boundaries, page layout information or graphics components, in order to either impose constraints for solving the system of linear equations, or to find vanishing points for homography estimation [22].

In [14], the authors discussed a competition (organized as part of ICDAR 2015) in order to evaluate the quality of mobile document capture and OCR. The competition consisted of two challenges, the Smartphone Document Capture and the Smartphone OCR. The first challenge was focused on page boundary detection, given a video clip, whereas participants of the second challenge had to extract textual content from document images. The evaluation dataset is based on the SmartDoc-QA dataset (see 4.1), which contains varying levels of perspective distortions.

In [18], the authors present a deep convolutional neural network for estimating the relative homography between a pair of images. Here, the main idea is to directly predict the homography H^{AB} that unwarps a distorted image B by using image A as a reference. However, instead of estimating the 3×3 matrix

itself, the 4-point parameterization is used. Given 4 corresponding points, $H_{3\times3}$ is computed by using e.g. the Direct Linear Transform (DLT) algorithm [19].

In [37], two vanishing points (horizontal and vertical) are extracted based on lines and line feeds, using the Radon transform, to estimate H.

In [9] the authors propose a convolutional neural network trained end-to-end on a large scale synthetic dataset. Their learning objective is to minimize the displacement vector from a canonical position.

3 Proposed Methods

Many recent approaches make use of manual feature engineering or rely on textual cues such as lines for homography recovery (e.g. [37]). However, these methods are typically not robust in uncontrolled environments, i.e. when applied to real world images [9]. Recently, deep learning techniques have outperformed traditional machine learning approaches in a wide variety of fields, such as Computer Vision, Natural Language Processing or Speech Recognition (e.g. [24]) and are therefore our method of choice.

3.1 Learning Objective

Our approach is based on the ideas presented in [9], which estimates the corner displacement vectors of the distorted document image from its canonical position. One key difference to the original work, though, is the way displacement vectors are learned when the document corners are outside of the image boundary:
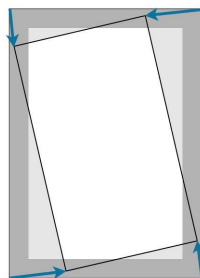
Figure 3: Learning of image corners outside the image boundary

In [9] the authors set image margins pixels to zero values (illustrated as gray margin), without changing the ground-truth annotations. However, this is different from the real scenario, where corner points may have negative coordinate values. Here, we take this scenario into account.

3.2 Network architectures

We explore two network architectures: a reimplementation based on the original work (referred to as NUCES-FAST in the following sections) and a modified version of the Xception-network [16], which recently outperformed the Inception-v3 architecture in terms of both the ImageNet ILSVRC [32] and the JFT [21] datasets. In both cases, the Adam optimization method [23] in combination with the L_1-loss is used:

$$L_1 = \frac{1}{N} \sum_{j=1}^{N} |(\boldsymbol{y_j} - \hat{\boldsymbol{y_j}})| \tag{1}$$

L_1 computes the mean absolute error between the ground truth ($\boldsymbol{y_j}$) and the predicted ($\hat{\boldsymbol{y_j}}$) corner points of the distorted document image j, over a training set of size N. Furthermore, the input image coordinates are normalized to $[0, 1]$.

3.2.1 NUCES-FAST

The first Neural network consists of the repetitive application of convolutional layers (filters of size 5x5 and 3x3), followed by a ReLU-activation function and max pooling (2x2). After the last convolutional layer, Dropout (0.5) is used as a regularization technique. Finally, a fully connected layer regresses the corner points of the distorted document image. The network has 11 convolution layers and a total of 34,115,464 trainable parameters, which are initialized using the Keras method he_normal [4].

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 384, 256, 64)	4864
max_pooling2d (MaxPooling2D)	(None, 192, 128, 64)	0
conv2d_1 (Conv2D)	(None, 192, 128, 128)	204928
conv2d_2 (Conv2D)	(None, 192, 128, 256)	295168
max_pooling2d_1 (MaxPooling2D)	(None, 96, 64, 256)	0
conv2d_3 (Conv2D)	(None, 96, 64, 384)	885120
conv2d_4 (Conv2D)	(None, 96, 64, 384)	1327488
max_pooling2d_2 (MaxPooling2D)	(None, 48, 32, 384)	0
conv2d_5 (Conv2D)	(None, 48, 32, 512)	1769984
conv2d_6 (Conv2D)	(None, 48, 32, 512)	2359808
max_pooling2d_3 (MaxPooling2D)	(None, 24, 16, 512)	0
conv2d_7 (Conv2D)	(None, 24, 16, 1024)	4719616
conv2d_8 (Conv2D)	(None, 24, 16, 1024)	9438208
max_pooling2d_4 (MaxPooling2D)	(None, 12, 8, 1024)	0
conv2d_9 (Conv2D)	(None, 12, 8, 1024)	9438208
conv2d_10 (Conv2D)	(None, 12, 8, 2048)	2099200
dropout (Dropout)	(None, 12, 8, 2048)	0
flatten (Flatten)	(None, 196608)	0
dense (Dense)	(None, 8)	1572872

Table 1: NUCES-FAST architecture

3.2.2 Xception

The Xception-network is based on the well-known Inception-network [36] and can be considered as a "linear stack of depthwise separable convolution layers with residual connections" [16]. According to the authors, the main motivation of the Inception-network is to simplify the convolution layer, because both cross-channel and spatial correlations are analyzed within one operation. Xception is an extreme version of it, since it entirely decouples this operation. Furthermore, residual connections are integrated into the network, inspired by the ResNet-architecture, which won the 1st place on the ILSVRC 2015 classification task [20]. In this work, the official implementation available on Github [15] is used – only the exit flow (see [16]) is slightly changed to regress the corner positions (see table 2). The network has 36 convolution layers and a total of 20,823,344 trainable parameters.

3.3 OCR Evaluation Framework

In order to enable comparisons across a wide range of datasets as well as metrics, we propose the following evaluation framework (see figure 4). The design decisions are inspired by the NILM-toolkit [10]. They enable computing both indirect (e.g. word/ character accuracy using the UNLV/ISRI Analytic Tools for OCR Evaluation [29]) and direct metrics (Multiscale Structural Similarity [39]). Currently, only Tesseract 4.0 is integrated into the evaluation framework. More tools like Ocropus3 [12] will be added in future releases.

To evaluate the performance of the perspective correction component, two experiments are performed: i) The mean character accuracy (in terms of edit operations [38] needed to correct the output) of Tesseract OCR, together with the associated confidence interval (95%), is determined based on the SmartDoc 2015 challenge 2 test set. No perspective correction is used here – this experiment serves as baseline. Note:

Layer (type)	Output Shape	Param #
block13_sepconv1_act (Activation)	(None, 24, 16, 728)	0
...	...	...
add_11 (Add)	(None, 12, 8, 1024)	0
block14_sepconv1 (SeparableConv2D)	(None, 12, 8, 1536)	1582080
block14_sepconv1_bn (BatchNorm)	(None, 12, 8, 1536)	6144
block14_sepconv1_act (Activation)	(None, 12, 8, 1536)	0
block14_sepconv2 (SeparableConv2D)	(None, 12, 8, 2048)	3159552
block14_sepconv2_bn (BatchNorm)	(None, 12, 8, 2048)	8192
block14_sepconv2_act (Activation)	(None, 12, 8, 2048)	0
avg_pool (GlobalAveragePooling2D)	(None, 2048)	0
predictions (Dense)	(None, 8)	16392

Table 2: Xception architecture with an exit flow modified for regression
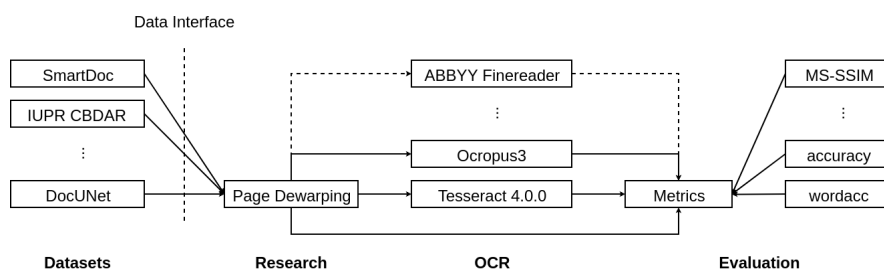
Figure 4: OCR evaluation framework

the official evaluation tools were provided by the organizers upon request. ii) The perspective correction component is plugged into the OCR pipeline before running Tesseract. The same metrics are computed as in experiment 1. Then the results are compared to the official results of the competition.

4 Data

Neural networks are data-driven models, which require large-scale training data. Since real data, together with the associated ground truth, is expensive to obtain, the training of the DNNs is mainly done on synthetic data. Therefore, the quality of the DNNs depends significantly on the quality of the data generator. This section reviews publicly available datasets for benchmarking and the procedure for generating a synthetic dataset.

4.1 Publicly available datasets

Publicly available datasets of camera-captured documents are summarized in table 3.

Dataset	Release	Institution	Size
DFKI-1 CBDAR	2007	DFKI	102
IUPR CBDAR	2011	DFKI	100
DIQD	2013	University of Maryland	175
SmartDoc-QA	2015	Université de La Rochelle	4260
DocUNet	2018	Stony Brook University, Megvii Inc.	130

Table 3: Available datasets of camera-captured documents

In [33] the authors of the Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) present the DFKI-1 Warped Documents Dataset in order to enable comparison of page dewarping techniques. A drawback of the DFKI-1 Warped Documents Dataset, however, is the lack of complex layouts and

sophisticated curl or skew [13]. Therefore, the IUPR CBDAR dataset of camera-captured document images was presented. It contains 100 grayscale document images from different book types (technical and non-technical) with diverse layouts as well as various perspective (camera viewing angle) and geometric distortions (folded/unfolded book). The ground-truth ASCII text and the corresponding flatbed scans are also provided.

In [25], the Document Image Quality Dataset (DIQD) is presented. The dataset contains a total of 175 document images captured from 25 physical documents of the same document type, with varying levels of blur, using a smartphone. Ground-truth text files are provided.

The SmartDoc-QA dataset provides a benchmark for quality assessment methods for mobile captured documents [30, p. 1]. The dataset has a total of 4260 document images captured from 30 different paper documents (modern documents, old administrative letters and receipts). The document images are captured under varying conditions (light, different types of blur and perspective angles) from two different mobile devices. The full text transcription of the documents is included as ground-truth.

The DocUNet-dataset offers a benchmark for the evaluation of the rectification process of distorted physical documents from a single image [27, p. 5]. It contains 130 original photos, document-centered cropped images, as well as scans from a flatbed scanner. The benchmark was created considering different document types (receipts, letters, academic papers, etc.), distortions (e.g. folded) and various environments (cameras, illumination conditions).

4.2 Synthetic Dataset Generation (DocHomography)

Our generator named DocHomography is based on the ideas presented in [9], where the generation of one training sample consists of the following steps. i) First, a document image is selected randomly from a set of images showing invoices. The target size of each image is set to (384×256), in order to enable comparison to [9]. ii) Second, a background image is selected randomly. Similar to [9], a subset of the MIT indoor scenes dataset [31] and the Describable Textures Dataset (DTD) [17] is used. iii) Next, the perspective distortion is augmented by randomly shifting the 4 corner points within a range of $[-s, s]$ pixels in xy-direction of the image plane, followed by computing and applying H (note: this is different from the original work, where the authors directly sampled different values of H coefficients from a uniform distribution). In this work, s is set to 100. iv) Finally, data augmentation techniques like Gaussian blur, additive Gaussian noise and gamma correction are used to improve the simulation of real images.
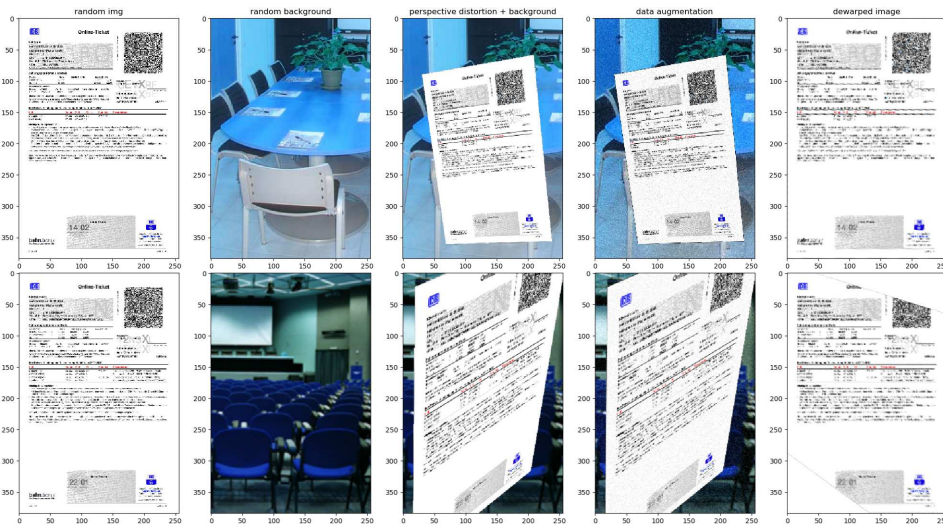

Figure 5: Examples of DocHomography generated images

The collection of training samples consists of images x together with their ground truth corner dis-

placement vectors $\boldsymbol{y}$. Fig. 5 illustrates the individual steps of the synthetic generation process. The images in the last column are obtained by perspectively correcting the distorted images using $\boldsymbol{y}$. In contrast to the original work [9], DocHomography allows to generate images where corners are outside of the image boundary (see figure 5, bottom row). The parameter p ($[0, 1]$) controls the probability of corner points being generated within the image boundary. For example, $p = 1$ means that all corner points are inside of the image boundary, whereas $p = 0.7$ allows negative displacement vectors with a probability of 30%.

Beside the possibility of generating an arbitrarily large dataset, DocHomography also allows to be used as online generator, where data is only generated batch-by-batch. This is particularly useful when the dataset is too big to fit into memory. For example, in order to train a model using an online generator, one can use the fit_generator method [3] provided by Keras.

5 Experiments and Results

This part evaluates the performance of the DNN trained to perspectively correct document images, using the framework described in section 3.3. First, the individual experiments are described (sections 5.1 and 5.2). Then we compare our results to those reported in the SmartDoc 2015 challenge 2 (section 5.3).

5.1 Experiment 1: Baseline

Tesseract 4.0 is used as an OCR engine with the following parameter settings [6]:

version	psm	oem	tessdata	lang
Tesseract 4.0.0-beta.1	3	1	tessdata_fast	eng

Table 4: Tesseract configuration

The last rows of tables 7 and 8 contain the Tesseract baseline results. In table 7, the OCR accuracy of Tesseract by itself is evaluated on the SmartDoc 2015 challenge 2 test set and compared to the competing methods. Table 8 presents results for various types and levels of distortion, also using the OCR accuracy as evaluation metric. In particular, we examined different levels of blur (no blur up to high blur), different light conditions (day time, multiple light sources, etc., referred to as L0 to L3) and different levels of perspective distortion (no distortion to extreme distortion). Those different conditions are described in more detail in [14].

Tesseract by itself achieves a character accuracy of 77.27%, which, on average, corresponds to 655 errors per page. The results in table 8 show that strong blurring effects (56.66%) have the biggest impact on the OCR accuracy, followed then by extreme perspective distortions (71.51%).

5.2 Experiment 2: Perspective Correction + Tesseract

In order to remove perspective distortion from camera-captured documents, we integrated our correction module into the OCR pipeline. We examined various network architectures and parameter settings to empirically determine a good solution. Again, we evaluated the extended pipeline on the SmartDoc 2015 challenge 2 test set.

5.2.1 Finding the best parameter setting

All experiments are run using tensorflow on a NVIDIA GEFORCE RTX$^{\text{TM}}$ 2070-GPU [2]. The experimental setup parameters are presented in table 5.

Epochs	Training size	Validation size	Batch Size
150	10,000	1000	16

Table 5: Parameter settings for the DNN training

Instead of using a pre-generated dataset, which is always limited by the disk space available, this work uses the online generator DocHomography, as described in 4.2. Here, a total of 1.5M unique training samples are used to learn the model's weights. Note: due to time and computing limits, the parameter values shown in table 5 are not further investigated but assumed to be reasonable. The experiments are run on RGB and grayscale images and consider both corner points inside and outside of the image boundary as described in section 4.2. The experimental results are summarized in table 6. Figure 6 illustrates the learning progress of the corresponding displacement vectors (loss in pixel) with respect to the number of epochs.

The following observations are made: i) The Xception architecture (validation error of 5.98px) outperforms the NUCES-FAST network (validation error of 9.67px) by almost 4 percentage points. In contrast to the NUCES-FAST network, the Xception network keeps improving with an increased number of epochs. A possible reason for this behavior might be the different network depth (Xception: 36 vs NUCES-FAST: 11) in conjunction with residual connections, which allow learning more complex and thus more powerful features. ii) By disallowing corner points outside the image boundary ($p = 1$), the Xception model achieves a validation error of 3.38px, which is comparable to the results presented in [9] (2.45px). iii) Unlike the observations made in [9], converting the images to grayscale does not harm the performance of our model.

id	architecture	color	p	training error	validation error
1	NUCES-FAST	rgb	0.7	9.486	9.67
2	Xception	rgb	0.7	6.35	5.98
3	Xception	gray	0.7	6.49	5.50
4	Xception	rgb	1	3.15	**3.38**

Table 6: Summary of trained models to recover homography
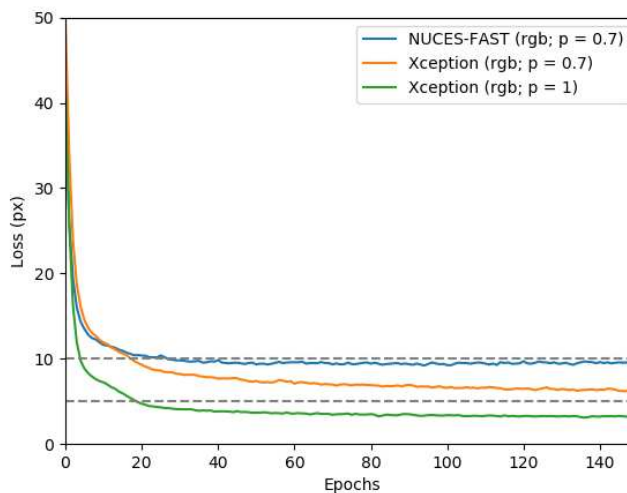

Figure 6: Visualization of the learning progress (loss in px) with respect to the number of epochs for different network architectures.

5.2.2 Plugging the homography model into the OCR Pipeline

The performance of our Xception model (rgb, $p=1$) as part of the OCR pipeline is shown in tables 7 and 8.
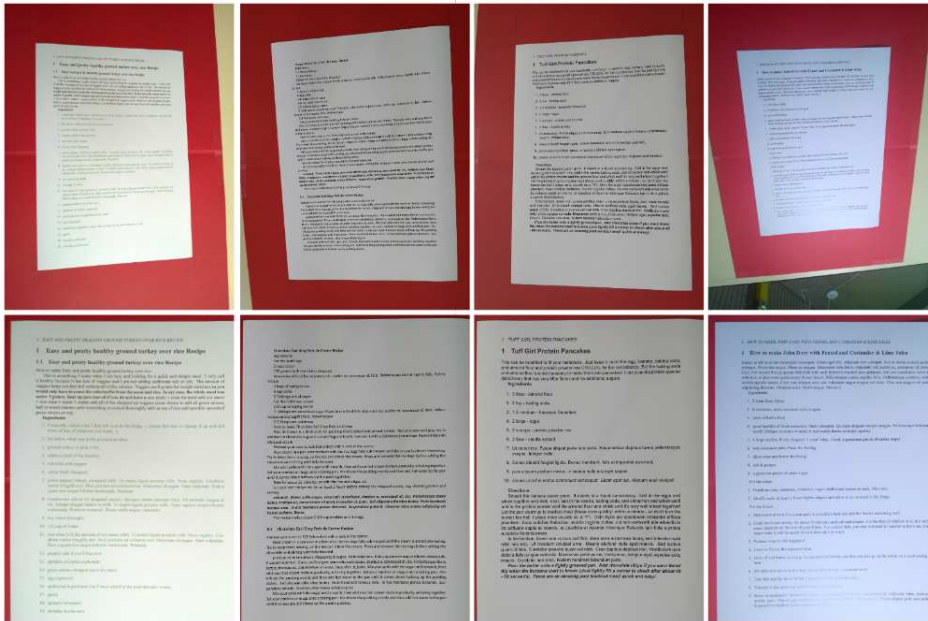
Figure 7: Examples of perspective correction: (top row) samples from the SmartDoc 2015 challenge 2 test set; (bottom row) corrected images.

Our extended pipeline achieves a character accuracy of 93.11%, which, on average, corresponds to 199 errors per page, i.e. 456 errors less in comparison to Tesseract by itself. We observe that the character accuracy i) decreases proportionally with increasing blur $(95.20 - 67.42)$ ii) is almost insensitive to different lighting (Standard Deviation: 1.94) iii) is almost insensitive to perspective distortions. The small accuracy drop (87.31) in the case where no perspective distortion is present is due to strong blurring effects and not to page dewarping failure.

Figure 7 shows some dewarping results. Figure 8 visualizes text lines recognized by Tesseract, indicated by green boxes, with and without dewarping. Since Tesseract was initially designed to rely on proper scans rather than on camera-captured document images, its page layout analysis algorithm [35] has obvious difficulties identifying text lines in case of perspective distortion.

5.3 Comparison to official results

This work's approach is now compared to the official results of the SmartDoc 2015 challenge 2 [5], with Tesseract 4.0 added as baseline. The competing methods (CCC, LRDE, etc.) are described in detail in [14].

Ranking	Method	acc (%)	conf. interval (%)	errors/page
1	CCC	99.93	[99.92,99.93]	2
2	LRDE	95.85	[95.56,96.14]	120
3	Digiform	95.33	[94.98,95.68]	135
4	A2iA	93.84	[93.54,94.15]	178
5	**This work**	93.11	[92.74 93.49]	199
6	CartPerk	91.19	[90.60,91.79]	254
7	FineReader 11	87.61	[87.20, 88.02]	357
8	Tesseract 4.0	77.27	[76.80, 77.74]	655

Table 7: SmartDoc 2015 challenge 2 comparison of competing methods

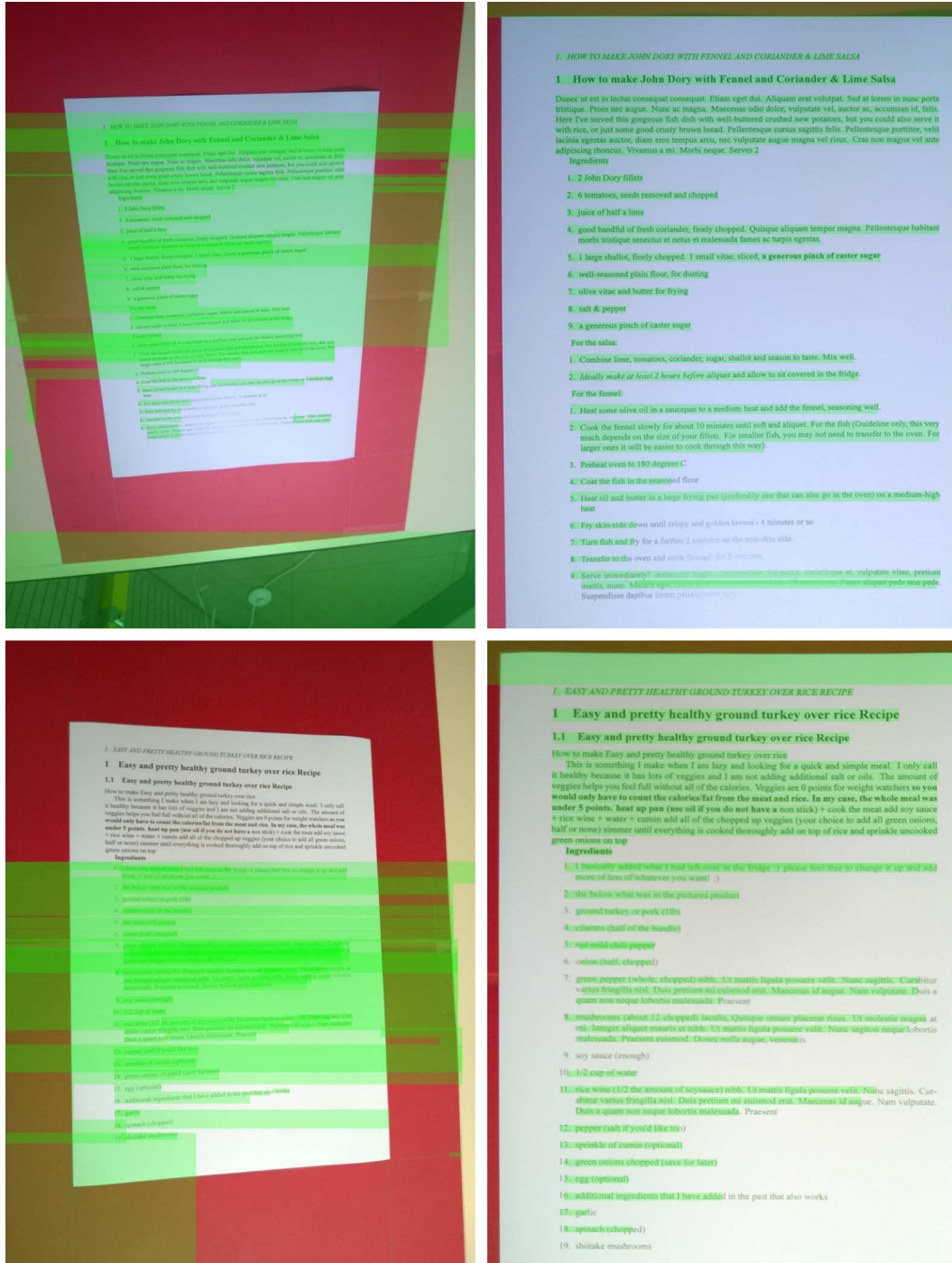The best result was achieved by the CCC-method (99.93% character accuracy), which won by a large

Figure 8: Text lines recognized by Tesseract, visualized as green rectangles, of samples from the Smart-Doc 2015 challenge 2 test set before (left column) and after (right column) dewarping.

margin (more than 5 percentage points). Unlike the other methods, CCC uses assumptions about the background color to detect and to dewarp the document image. In addition, CCC does training both on sharp and blurry grayscale text lines which leads to significantly improved performance on blurred images.

This work's approach is placed fifth, with a character accuracy of 93.11%. It improves Tesseract 4.0 (by itself) by more than 15 percentage points, and it outperforms the results of an earlier version of the commercial OCR solution ABBYY FineReader [1] (table 7).

Method (acc)	blur				light				perspective		
	no	low	med	high	L0	L1	L2	L3	no	med	extrm
CCC	99.93	99.93	99.92	99.93	99.93	99.93	99.93	99.93	99.92	99.93	99.93
LRDE	96.65	97.82	91.68	83.36	94.6	95.34	97.11	96.43	93.88	96.37	95.23
Digiform	96.73	97.13	88.71	74.47	93.63	94.01	96.83	97.1	90.98	95.76	96.35
A2iA	95.46	91.41	82.79	77.19	92.26	94.71	94.52	93.79	89.56	94.94	92.65
This work	95.20	93.01	80.38	67.42	89.77	94.68	93.89	94.01	87.31	94.31	92.44
CartPerk	93.95	94	70.89	57.66	87.01	90.58	94.03	93.34	82.76	92.5	91.59
FineReader 11	-	-	-	-	-	-	-	-	-	-	-
Tesseract 4.0	79.32	73.74	63.17	56.66	76.99	76.97	77.92	77.23	78.85	78.76	71.51

Table 8: SmartDoc 2015 challenge 2 comparison of competing methods on distorted images

The method is robust against a variety of different lighting and perspective distortion, but not against blurring effects (see table 8). The OCR accuracy could be improved by either fine-tuning or retraining the weights of a pretrained language model [8] on the SmartDoc dataset.

Finally, it should be noted that the Xception model is not only accurate but also suitable for real-time applications [11] with little additional time overhead compared to Tesseract alone.

6 Conclusion and Future Work

In this work, Tesseract has been extended by a perspective correction component, such that it can be used as a mobile OCR scanner. The correction method assumes that a document lies in a plane and estimates the corner point positions to recover the desired homography. Experiments have shown that the OCR pipeline produces reliable results on the SmartDoc 2015 challenge 2 test set. However, it remains uncertain if the correction component generalizes well to different datasets and therefore requires further investigations.

In future work, the goal is to integrate an unrestricted page dewarping component suitable for processing both curved and folded document images (including perspective distortions) into the OCR pipeline. For example, our data generator could be extended to synthesize folded and crumpled documents as described in [28]. However, this is not a trivial step and would probably require much more training data. Furthermore, the OCR evaluation framework (3.3) could be extended to provide a uniform basis of comparisons.

References

[1] ABBYY FineReader 11. https://www.abbyy.com/en-ee/support/finereader-11/. [Online; accessed 16-January-2019].

[2] GeForce RTX 2070. https://www.nvidia.com/de-de/geforce/graphics-cards/rtx-2070/. [Online; accessed 04-October-2018].

[3] Keras: The Sequential model API. https://keras.io/models/sequential/. [Online; accessed 11-October-2018].

[4] Keras: Usage of initializers. https://keras.io/initializers/. [Online; accessed 07-October-2018].

[5] Results on SmarDoc 2015 – challenge 2. http://smartdoc.univ-lr.fr/results-report/. [Online; accessed 03-January-2019].

[6] Tesseract CLI. https://github.com/tesseract-ocr/tesseract/wiki/Command-Line-Usage. [Online; accessed 12-November-2018].

[7] Tesseract Open Source OCR Engine. https://github.com/tesseract-ocr/tesseract. [Online; accessed 26-November-2018].

[8] Training Tesseract 4.00. https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract-4.00. [Online; accessed 20-December-2018].

[9] S. A. Abbas and S. ul Hussain. Recovering homography from camera captured documents using convolutional neural networks. *CoRR*, abs/1709.03524, 2017.

[10] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. Nilmtk: An open source toolkit for non-intrusive load monitoring. 07 2014.

[11] S. Bianco, R. Cadène, L. Celona, and P. Napoletano. Benchmark Analysis of Representative Deep Neural Network Architectures. *CoRR*, abs/1810.00736, 2018.

[12] T. M. Breuel. OCRopus3. https://github.com/NVlabs/ocropus3. [Online; accessed 12-November-2018].

[13] S. S. Bukhari, F. Shafait, and T. M. Breuel. The IUPR Dataset of camera-captured document images. In *Camera-Based Document Analysis and Recognition*. Springer Berlin Heidelberg, 2012.

[14] J. Burie, J. Chazalon, M. Coustaty, S. Eskenazi, M. M. Luqman, M. Mehri, N. Nayef, J. Ogier, S. Prum, and M. Rusiñol. ICDAR2015 competition on smartphone document capture and OCR (SmartDoc). In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1161–1165, 2015.

[15] F. Chollet. Trained image classification models. https://github.com/fchollet/deep-learning-models. [Online; accessed 22-November-2018].

[16] F. Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016.

[17] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. Describing Textures in the Wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[18] D. DeTone, T. Malisiewicz, and A. Rabinovich. Deep image homography estimation. *CoRR*, abs/1606.03798, 2016.

[19] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[20] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[21] G. Hinton, O. Vinyals, and J. Dean. Distilling the Knowledge in a Neural Network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[22] L. Jagannathan and C. V. Jawahar. Perspective correction methods for camera-based document analysis. In *Internation Workshop on Camera-Based Document Analysis and Recognition*, pages 148–154, 2005.

[23] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[25] J. Kumar, P. Ye, and D. Doermann. A dataset for quality assessment of camera captured document images. In *Camera-Based Document Analysis and Recognition*. Springer International Publishing, 2014.

[26] N. Körber. GitHub repositories. https://github.com/Nikolai10?tab=repositories. [Online; accessed 14-January-2019].

[27] K. Ma, Z. Shu, X. Bai, J. Wang, and D. Samaras. Docunet: Document image unwarping via a

stacked u-net. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[28] R. Narain, T. Pfaff, and J. F. O'Brien. Folding and crumpling adaptive sheets. *ACM Transactions on Graphics*, 32(4):51:1–8, July 2013. Proceedings of ACM SIGGRAPH 2013, Anaheim.

[29] T. A. Nartker, S. V. Rice, and S. E. Lumos. Software tools and test data for research and testing of page-reading OCR systems. In *Proceedings of SPIE - The International Society for Optical Engineering*, pages 37–47, 2005.

[30] N. Nayef, M. M. Luqman, S. Prum, S. Eskenazi, J. Chazalon, and J.-M. Ogier. Smartdoc-QA: A dataset for quality assessment of smartphone captured document images - single and multiple distortions. In *Proceedings of the sixth international workshop on Camera Based Document Analysis and Recognition (CBDAR), 2015*.

[31] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 413–420, June 2009.

[32] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F. F. Li. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115, 09 2014.

[33] F. Shafait and T. M. Breuel. Document image dewarping contest. In *in 2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, pages 181–188, 2007.

[34] R. Smith. An overview of the tesseract OCR engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633, Sep. 2007.

[35] R. Smith. Hybrid page layout analysis via tab-stop detection. In *Proceedings of the 10th international conference on document analysis and recognition*, 2009.

[36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[37] Y. Takezawa, M. Hasegawa, and S. Tabbone. Robust perspective rectification of camera-captured document images. In *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017.

[38] R. A. Wagner and M. J. Fischer. The String-to-String Correction Problem. *J. ACM*, 21(1):168–173, Jan. 1974.

[39] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.

Nikolai KÖRBER
University of Applied Sciences Landshut
Faculty of Computer Science
Am Lurzenhof 1, 84036 Landshut
GERMANY
E-mail: *nikkoerber@gmail.com*

Enhancements in Human-Computer Interaction

Adrian Pintilie

Abstract

Despite the fact that we witness more and more breakthroughs in technology each year, the interaction between humans and computers is still limited to pushing buttons or moving the mouse, as even the most advanced virtual assistants available (e.g. Siri, Cortana) are able to perform only very basic tasks. The current work outlines an interface that aims to make the human-computer interaction more humanlike, using only video and audio input. Powered by state-of-the-art technologies, the interface provides control over a variety of functions, from basic tasks like moving the cursor or changing the sound volume, to more complex tasks such as playing video games or browsing the internet.

1 Introduction

To make the interaction between humans and computers more humanlike it is required to focus our attention on video and audio input. To achieve this, the first step is making a computer to see, and extract information from what it sees, to hear, and extract information from what it hears. The second step is making the computer capable of performing complex actions based on the acquired information.

Since 2012, when the paper "ImageNet Classification with Deep Convolutional Networks" [1] was published, widely regarded as one of the most influential publications in the Computer Vision field, extracting information from images became much easier and efficient. The development of these computing systems made possible very complex tasks like object segmentation or object detection. Although these new and promising technologies were adopted in many fields, no applications were developed for enhancing the human-computer interaction.

The advancements in Natural Language Processing allowed virtual assistants to handle seamlessly difficult tasks such as speech recognition, natural language understanding, and natural language generation. However, the virtual assistants available on desktop systems are limited to tasks like scheduling an alarm or answering a limited set of questions.

Therefore, using state-of-the-art technologies to achieve the above-mentioned first step, and innovative approaches to achieve the above-mentioned second step, the project aims to provide an interface that makes the human-computer interaction more intuitive and easier to use, an interface that unbounds the user from the hardware, in other words, an interface that is more humanlike.

The present paper is structured in sections. The first section will focus on enhancements in human-computer interaction using video input, the second section will focus on enhancements in human-computer interaction using audio input, followed by a final section that will present the conclusions and future work.

2   Enhancing Human-Computer Interaction using video input

2.1 Building the necessary infrastructure to collect information from video input

As mentioned above, 2012 marked the Computer Vision field with the power of Convolutional Neural Networks. Based on supervised learning, these types of deep neural networks are specially designed to process pixel data. Taking as input an image, CNN-s are able to classify the image to a specific category.

As the traditional CNN-s are only capable of classifying images, that is, for each input image the result is a single predicted category, there are a number of limitations that emerge. The first scenario that worth mentioning is the scenario where in the same image we have multiple objects that belong to different categories. Because the output of a traditional CNN is a single category, all the information about the remaining objects is lost (e.g. if we have an image with two hands and one head, the output might be "hand" or "head"). Another important limitation is the fact the output lacks information about the localization of the object in the image. In order to provide extensive control over a variety of functions, information on both the number of detected objects and their localization is required.

Object detection is a computer technology that solves these problems by performing object instance segmentation. This technology uses Region Based Convolutional Neural Networks, and the first version of this architecture, called R-CNN [2], was published in 2014. Since then, three more published papers demonstrated improved architectures in terms of speed and accuracy: Fast R-CNN in 2015 [3], Faster R-CNN in 2016 [4] and Mask R-CNN in 2018 [5]. Some may argue that the advent of R-CNNs has been more impactful than any of the already existing papers in the Computer Vision field.

I opted for the Faster R-CNN architecture due to the fact that it represents the optimal trade-off between speed and accuracy. Since the most suitable architecture was identified, the next step is training a neural network with a dataset that meets our necessities.

Sign languages are languages that use visual-manual modality to convey meaning. Since these languages were proven efficient, a similar system can be used to map a specific set of hand signals to suitable actions. A number of comprehensive sign-language datasets are available online. However, these datasets are suitable to train only traditional CNNs. For both traditional CNNs and R-CNNs labeled data is required, the difference is that for the first category we have only one label for each observation. R-CNNs have to two general components, the region proposal step, that is a region proposal network (RPN) for generating region proposals by selective search, and the classification step, that is a network using the proposals to detect objects. As a result, the training data has to have a different format for this architecture: for every object from all the images, it is required to have a labeled bounding box. To prepare the dataset, I used LabelImg, a graphical image annotation tool.
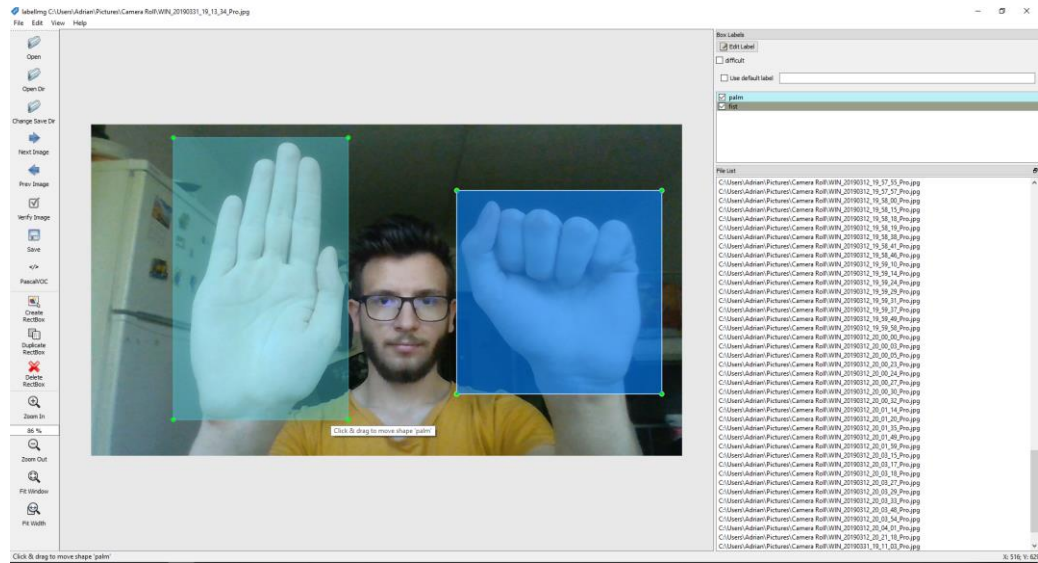
Figure 1: The annotation process using LabelImg

As expected, the annotation process is very costly. In addition, most models that solve complex problems require huge amounts of data in order to achieve good results. To minimize the duration of this process I decided to use transfer learning.

Transfer learning is a research problem that aims to save information gained while solving one problem and using this information to solve related problems. In this case, I used as starting point the weights of a model trained on the Common Objects in Context dataset (a dataset with 80 classes and 120.000 images, designed for object detection) to train the model with my own dataset, fine-tuning the layers from the pre-trained model.

For this model, to improve the speed, I chose to use for the classification step a CNN architecture called GoogleLeNet [6]. GoogleLeNet devised a module called Inception module that approximates a sparse CNN with a normal dense construction, and also, a so-called bottleneck layer that helps in the massive reduction of the computation requirements. Further, using a momentum optimizer with an initial learning rate of 0.0002, scheduled to decrease as the number of steps increase, and tweaking more hyperparameters, I was able to bring the value of the total loss below 0.1, for a dataset of approximately 2000 images with 7 classes. Monitoring the evolution of the training process using TensorBoard, transfer learning allowed me this way to achieve state-of-the-art accuracy with an impressive number of almost 300 images for each class.
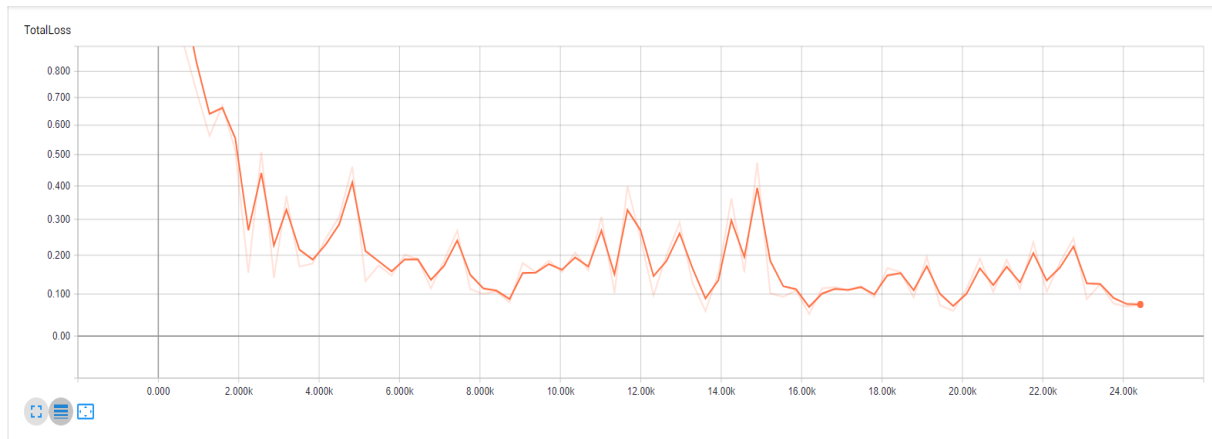
Figure 1: The evolution of the total loss

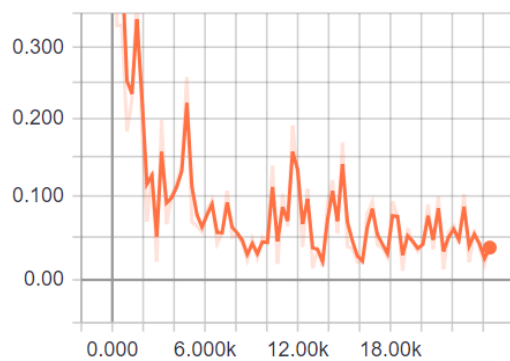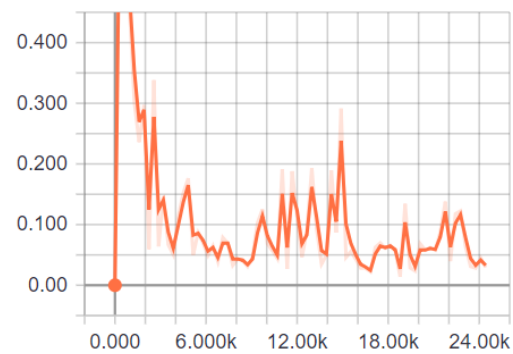
Figure 2: The evolution of
the classification loss

Figure 3: The evolution of
the localization loss

To improve the speed of the neural network, I used the NVIDIA CUDA Deep Neural Network library (cuDNN), a GPU-accelerated library of primitives for deep neural networks. This way, I managed to achieve object detection at a speed of approximately 7 frames per second on an NVIDIA GTX 1050 GPU.
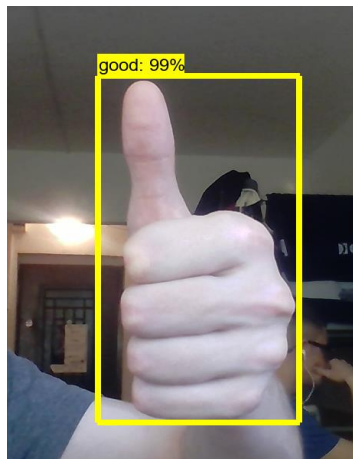
Figure 4: Detection of a fist with
a thumb up with 99% confidence
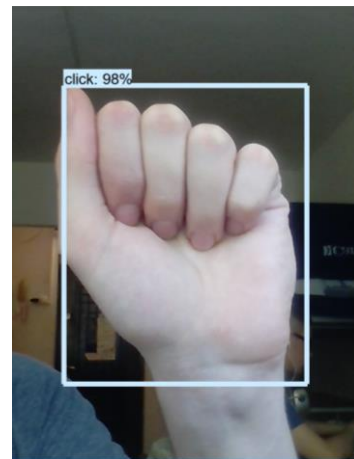
Figure 5: Detection of a palm
with 98% confidence

Real-time object detection provides all the necessary features to be able to handle user instructions.

2.2 Controlling the sound volume

Control over the sound volume is an important feature for all the devices that incorporate speakers. For laptops and desktop systems, the control of this characteristic is bounded to pushing buttons or scrolling, for TVs or audio systems is bounded to the use of remote controls. One way to facilitate access is to have several signs that a computer can interpret as signs for the manipulation of this characteristic.

To implement this functionality, I choose two intuitive signs, one labeled as "vol_up" to increase the sound volume, and one labeled as "vol_down" to decrease the sound volume.
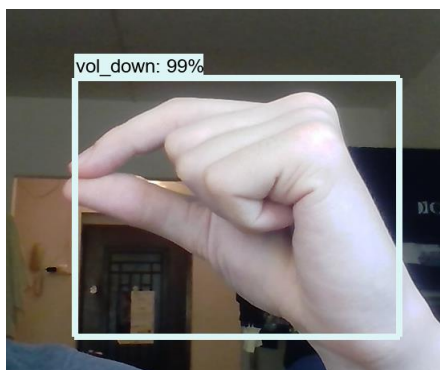
Figure 6: Detection of the "vol_down" class
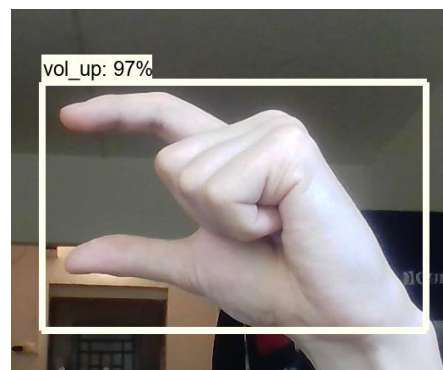
Figure 7: Detection of the "vol_up" class

Once one of these classes is detected, using the current value of the master sound volume, I am able to increase or decrease the value of this characteristic.

The code routine that handles the changes for the sound volume:

```
class SpeakerController:
  change_rate = 1.2

  def __init__(self):
    self.devices = AudioUtilities.GetSpeakers()
    self.interface = self.devices.Activate(IAudioEndpointVolume._iid_, CLSCTX_ALL, None)
    self.volume = cast(self.interface, POINTER(IAudioEndpointVolume))
    self.volume_lower_bound, _, self.volume_upper_bound = self.volume.GetVolumeRange()
    self.volume_change = self.volume_lower_bound * SpeakerController.change_rate / 100

  def increase_volume(self):
    current_volume = self.volume.GetMasterVolumeLevel()
    new_volume = current_volume + self.volume_change

    self.volume.SetMasterVolumeLevel(min(new_volume, self.volume_upper_bound), None)

  def decrease_volume(self):
    current_volume = self.volume.GetMasterVolumeLevel()
    new_volume = current_volume - self.volume_change

    self.volume.SetMasterVolumeLevel(max(new_volume, self.volume_lower_bound ), None)
```

2.3 Controlling the cursor

Nowadays, the mouse and the keyboard represent the most important parts in the interaction with a computer. To be able to control the cursor covers most of the functions required to interact with a computer, but the movement of the cursor is tied up to a mouse or a touchpad. This means having to deal with wires or batteries in the case of the mouse, or being stuck near a computer to be able to use it in the case of the touchpad.

An alternative would be to use a sign that will be interpreted as the cursor. If the position of the sign is known, we can use this information to manipulate the position of the cursor. One obvious approach could be to map the coordinates of the center of the sign to a certain position of the cursor on the screen, but this approach has a number of issues of which I will mention:

1. The screen resolution is usually much higher than the resolution of the image captured using a web camera. The higher the resolution of the screen, the lower will be the precision of the mapping.
2. As the user gets further from the camera, more ample movements will be required to reach certain parts of the screen, because the scale of the subject decreases as the distance increases.

A solution that solves these issues is to implement a system that moves the cursor based on the position of the sign relative to the center of the image.

The algorithm will work as follows:
1. The center of the sign is computed when one is detected
2. Knowing the coordinates of the center of the image and of the center of the object we are able to compute the slope and the distance between these two points.
3. Knowing the distance and the slope, we are able to project the new position of the cursor relative to its old position. The cursor will be moved relative to its old position a distance that is directly proportional to the distance between the center of the image and the center of the

sign in the image, and the direction is determined using the slope between the center of the image and the center of the sign in the image.

This way, the user will be able to perform slow and accurate movements when the position of his hand is near the center of the image, and more rapid movements as his hand gets further from the center of the image. Using this algorithm, support for precise and fluid movements of the cursor is provided.

Further, as Faster R-CNN returns the coordinates of the bounding boxes of the detected objects, the area of the object can be computed. Using the area of the object we can determine the distance of the user from the camera. Multiplying the distance used to move the cursor by a factor that is inversely proportional to the area of the detected object allows for precise movements even when the user gets further from the camera, solving this way the issue with the number two.
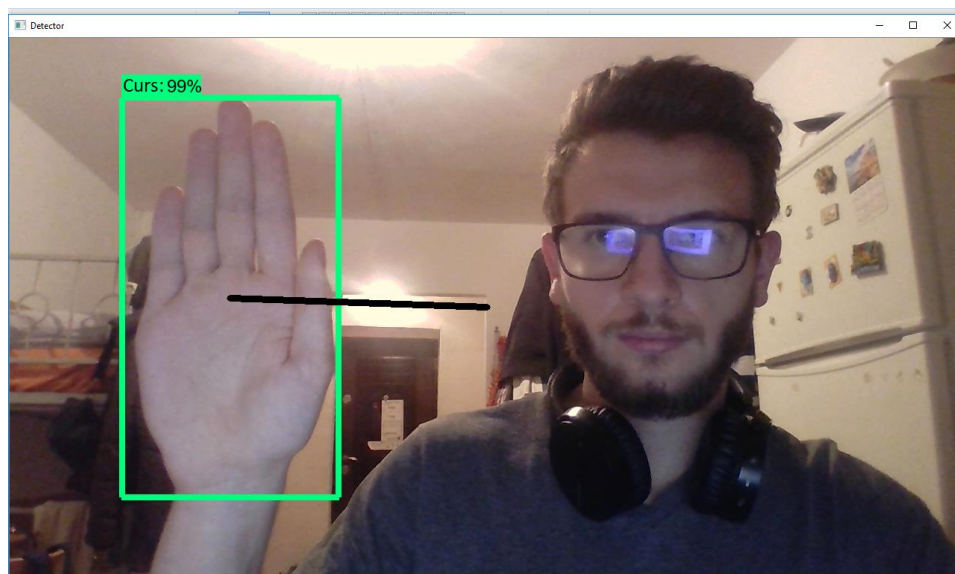

Figure 8: Detection of the "curs" class used for moving the cursor. Lines are drawn in real time to emphasize the position of the center of the object relative to the center of the image.

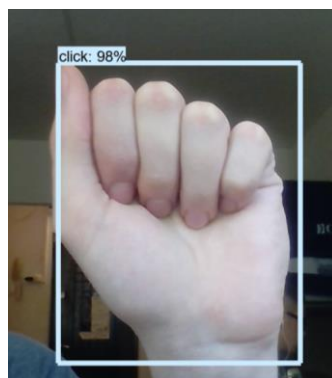After the user brings the cursor at the desired position, he is able to click by clenching his fist.

Figure 9: Detection of the "click" class.

2.4 Controlling video games

The gaming industry has evolved up to the point of photorealistic games, but let me take as an example the racing video games: playing racing video games using keystrokes is not even close to the real experience of driving, while gaming steering wheels represent expensive hardware.

An alternative would be an interface that is able to interpret gestures made by the user that are similar to the real experience of driving. The interface that I developed is based on an imaginary steering wheel: if the user raises two fists in the front of the camera, the interface knows to interpret these fists as a steering wheel. Having two fists allows the computation of the slope between the centers of these two objects, and as a result, allows the simulation of a real steering wheel, getting closer to the real driving experience.

The real world steering wheel controls the direction of the car in concordance with two factors: the angle of the wheel and the amount of time the wheel is fixed in that position. As the racing video games take as input keystrokes, the direction of the virtual car is controlled by two similar factors: a key that represents the direction (i.e., left or right), and the amount of time the key is fixed in that position. In other words, if we would use the same system that is used in video games on a real-world steering wheel, we would end up having a car that would not care about how much you steer the wheel, and care only about if the wheel is turned to left or right or not turned at all, and how much the steering wheel is held pointing to one of this direction. Therefore, the main challenge is converting a system that considers only if a slope is positive or negative and the amount of time the slope is positive or negative, to a system that considers also, how positive or how negative the slope is.

The solution to this problem was taking the value of the slope, normalize the value, and at each frame start a thread that sends a signal for a period of time that is directly proportional to the value of the slope.
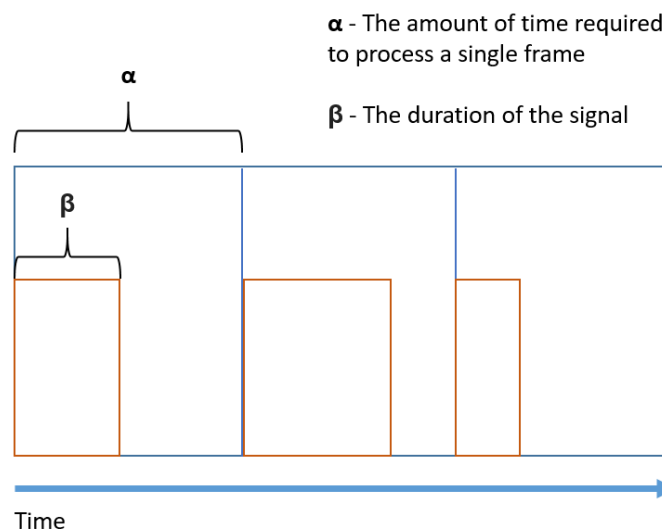
Figure 10: Visual representation of the steering system implementation
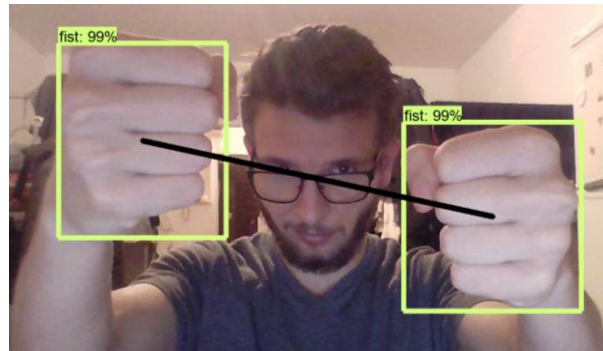
Figure 11: Visual representation of the steering system. Lines are
drawn in real-time to emphasize the position of the steering wheel.

At this point, a steering system is implemented. Next, support for acceleration and a braking system is required. To improve the user experience, both functionalities should be implemented by making use of hand gestures. In addition, to get closer to the real experience of driving, the interface should be able to distinguish between three scenarios: just steering, steering while accelerating and steering while braking.

To meet both these requirements, for acceleration and steering, I used a different class: when the interface detects two hands with the thumbs up, accelerates the virtual car, continuing to keep track of the slope between these two objects. This is the same algorithm used only for steering, but this time, a signal for acceleration is sent continuously.
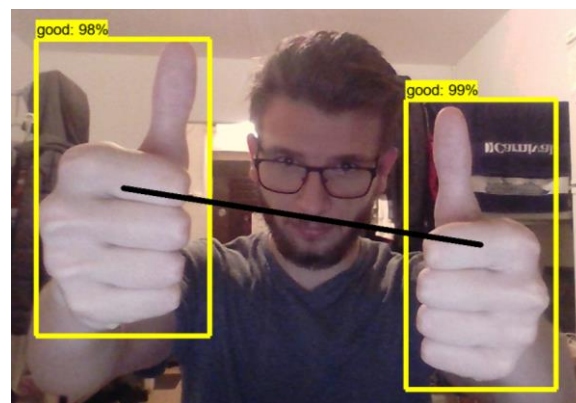
Figure 12: Visual representation of the steering system with support for
acceleration. Lines drawn in real-time to emphasize the position of the steering wheel.

To meet these requirements for braking and steering, I used the original class used for steering, but added an additional part that computes the area of the detected objects. This way, if the user is moving away his hands for a considerable distance, that is, the area of the detected objects decreases significantly, the interface interprets the gesture as pressing the brake, while still keeping track of the slope between these two objects.

3   Enhancing Human-Computer Interaction using audio input

3.1 Building the necessary infrastructure to collect information from audio input

To be able to perform complex tasks like browsing the internet based on voice commands from the user, the first step is to extract the information from the audio input. Speech recognition is a very complicated task, and as a result, to obtain state-of-the-art results I used Google's API that provides speech-to-text conversion powered by machine learning. This way, to avoid the permanent recording, after the user makes a specific gesture, his audio instruction is recorded and then processed by Google's API that returns the text version of the instruction. The main advantage of using this API is the possibility of using multiple languages. After the information from the audio input was extracted, the next step is to use this information to control the browser. To have extensive control over the browser I used Selenium, a portable framework for testing web applications.

3.2 Using voice commands for controlling the browser

Browsing the internet is probably the most popular activity on a computer and this is the reason why support for a wide range of queries should be added. The problem is that the websites are complex, and each one of them comes in a different format. A solution to this problem is to provide enhanced support for the most popular platforms, such as Google or YouTube, and a unified way of access for all the other websites.

To be able to search something on Google, there are two steps, the execution of a query and the selection of one of the results. For the first step, the instruction of searching on Google has the format "Search on Google", followed by the content that the user wants to search. Because the instruction starts with "Search on Google", using Selenium, I change the webpage to Google's home page. Here, because the format of this webpage is known, by parsing the webpage to find its main specific search box, we are able to send the content of the user. For the second step, the instruction of selecting one of the results has the format "Select", followed by a part of the title of the desired result. To make the "Select" instruction more generic, when the instruction is used, we start by parsing the current URL to identify the current website. In this case, the resulted website is Google, and because the format of Google's search page is known, the content of the webpage is parsed to find the titles of all the resulted websites. The titles are stored in an array, and the selected link will be the first link that contains the content of the user. The behaviour is similar for every platform based on similar usage, that is, searching and selecting, such as YouTube or Wikipedia. The main difference is that specific implementation for the format of every webpage has to be provided.

To be able to interact with complex websites that do not have complete support, after the link was selected from the Google's search page, the user still has to use the cursor, but to avoid using the keyboard the user can use the instruction "Navigator type" followed by the content to be typed. This functionality is useful for scenarios like sending messages using Facebook's Messenger.

Support for functions such as going back, going forward, or refreshing the webpage, was added using simple JavaScript functions.

3   Conclusions and future work

Considering the results mentioned above, this paper successfully presented the implementation of features that make the human-computer interaction more humanlike, exceeding, in terms of provided features, all the available commercial solutions.

Although the current interface handles proficiently voice instructions that have a fixed format, an important improvement would be to make the interface capable of processing natural language. Moreover, making the web browser completely independent of the cursor would mean a valuable alternative for people with motor disabilities.

With the current technologies available, this paper represents a starting point for enhancements in Human-Computer Interaction, where the only limit is our imagination.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*. In Advances in neural information processing systems, 2012.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich, *Feature hierarchies for accurate object detection and semantic segmentation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[3] R. Girshick, *Fast R-CNN*. In IEEE International Conference on Computer Vision (ICCV), 2015.

[4] S. Ren, K. He, R. Girshick, and J. Sun, *Faster R-CNN: Towards real-time object detection with region proposal networks*. In NIPS, 2015.

[5] K. He, G. Gkioxari, P. Dollár, R. Girshick, *Mask R-CNN*. In IEEE International Conference on Computer Vision (ICCV), 2017.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] Koen E. A. van de Sande, Jasper R. R. Uijlings, Theo Gevers, and Arnold W. M. Smeulders, *Segmentation as selective search for object recognition.* In International Conference on Computer Vision (ICCV), 2011.

Adrian PINTILIE
Transilvania University of Brașov
Faculty of Electrical Engineering and Computer Science
Bulevardul Eroilor 29, Brașov 500036
ROMANIA
E-mail: pintilie_adrian29@yahoo.com

The formation of a set of signs for modeling the evaluation of enterprise activities

Lyudmila Sabadash, Pavlo Grynko

Abstract

The article presents the results of solving the problem of formation of a sign space for the development of models of enterprise performance evaluation. This problem is complicated. Actuality of the solution of this problem is conditioned by the need to increase the statistical quality of the developed models of enterprise performance evaluation. In the mathematical literature there are no recommendations for constructing such a space of signs using mathematical methods. The purpose of this article is to describe the results of solving the problem of formation of a representative modeling space for evaluating the enterprise. The effectiveness of the mathematical model depends on the correct conceptual model of the enterprise. The authors recommend the simulation of the evaluation of the enterprise to carry out in its sign space, which consists of elementary, complex and latent signs of activity, which are obtained using mathematical methods and models. To determine causation relationships that reflect latent factors and form mechanisms in the economy, it is advisable to use factor and canonical analysis that are implemented in the software environments MatLab, Statgraphics, Statistica. The authors recommend that the hallmark of the model of enterprise performance evaluation be used as information for the development of managerial decisions.

Introduction

To organize the evaluating procedures the company's activities, it is important to correctly develop its mathematical model, which in turn depends on the features system that reflect this activity.

The constructed model only has a high statistical quality if it was formed in a qualitative sign space. The analysis of the scientists work allowed to generalize the requirements, which are advanced to the sign space formation [1- 6]. The content of these requirements: 1) the features should reflect the main properties of the modeling object; 2) the list of features should reflect the conceptual nature of the object; 3) the attributes system is hierarchical and consists of elementary and complex signs; 4) indicators and criteria should be distinguished among the signs; 5) signs are measured in metric and nonmetric scales, which determines the choice of the mathematical method for object modeling, and are measures of enterprise evaluation; 6) signs can be determined in conditions of certainty and uncertainty; 7) for the further modeling of the sign, the descriptive statistics methods should be used, putting in other words, to conduct an exploratory analysis; 8) the relationship between the features of the object in economy reflects causal mechanisms that are basis of its life, development and management; 9) the attributes are factors that allow determining

the key evaluation and the results factors; 10) signs are obvious and latent, generalizing and integral. In the mathematical literature there are no recommendations for constructing such a space of signs using mathematical methods. The purpose of this article is to describe the results of solving the problem of formation of a representative modeling space for evaluating the enterprise.

Methodology of research

Assessment of an enterprise modeling should be based on its multicriteria. The evaluation criterion is a measure that is defined on different scales to ensure reliability, adequacy and relevance to objective reality. Such definition is consistent with the understanding of the criterion in various spheres of human activity. Another matter is how this criterion will be measured, what tools and how to display them in models. So, for the most part, criteria for the development of economical and mathematical models are indications, both elementary and complex, which are measured by metric scales.

So you say, the developmen criteria of economical and mathematical models are complex and elemental features that reflect the properties, that is, the object characteristics, and express the quality and quantity for the object evaluation. In economic-mathematical modeling, there are general and partial criteria. The general criterion is usually constructed and is a function in economic-mathematical modeling, that is, served as a target function that sets the quality level of a object complex feature and can be investigated for optimum. The general criterion can be a vector function, the concept of which is a generalization of the function concept. Partial criteria are used to evaluate the complex objects characteristics in economy [5].

The problem understanding of multicriteria in the economy somewhat contradicts the statement that in practice multicriterial problems arise when it is possible to formulate and formalize in the form of criteria only a number of separate requirements that are put forward for an optimal solution and to combine these separate criteria into one does not have opportunities. Such a statement is valid in the mathematical objects modeling in technology, in economy of the other business. As a rule, in economy, the general criterion is known, as well as known and partial criteria, which should be taken into account in order to achieve full-scale, complexity, systematic representation of the properties of the object. The use of criteria in the economy is explained by the need to present, main features that determine the object content and the neglect of secondary, which only complement the basic.

Results

It should be noted that the effectiveness of the mathematical model depends on the correct conceptual enterprise model. The economic activity of industrial enterprises consists of the main activity types, namely financial, industrial, marketing, personnel, export-import, innovation-investment, which reflect the economic, production, organizational, economic and internal economic and external relations of a modern large industrial enterprise. The modeling practice estimation of the activity of an industrial enterprise shows that the influence component of the external environment has a significant relationship with other components that characterize the enterprises activity. Consequently, the conceptual model of the activity of a large industrial enterprise for its evaluation in the general form should be represented as follows:

$$IEA = \langle F, P, M, H, II, EI, IEE \rangle,$$

where IEA - industrial enterprise`s activity, F - financial activity, B - production activity, M - marketing activity, H – human resorces activity, II - innovation and investment activity, EI - export-import activity, IEE - impact of external environment.

In turn, the complexity, full scale and adequacy of the industrial enterprise assessment is provided by a hierarchical system of indicators, which contains partial (physical quantities and coefficients) and integral indicators that determine the elementary and complex characteristics of the activity [7]. Thanks to the system of partial indicators, it is possible to determine in detail the basic characteristics of the enterprise's activity, and determining the level of its state and development is based on the integral indicator value. In the generalized form, the hierarchical structure of the indicators system for the enterprise assessment is presented in fig.1.

Evaluation modeling of an enterprise's activity is carried out in its sign space, which consists of elementary, complex and latent signs of activity and obtained using mathematical methods and models.
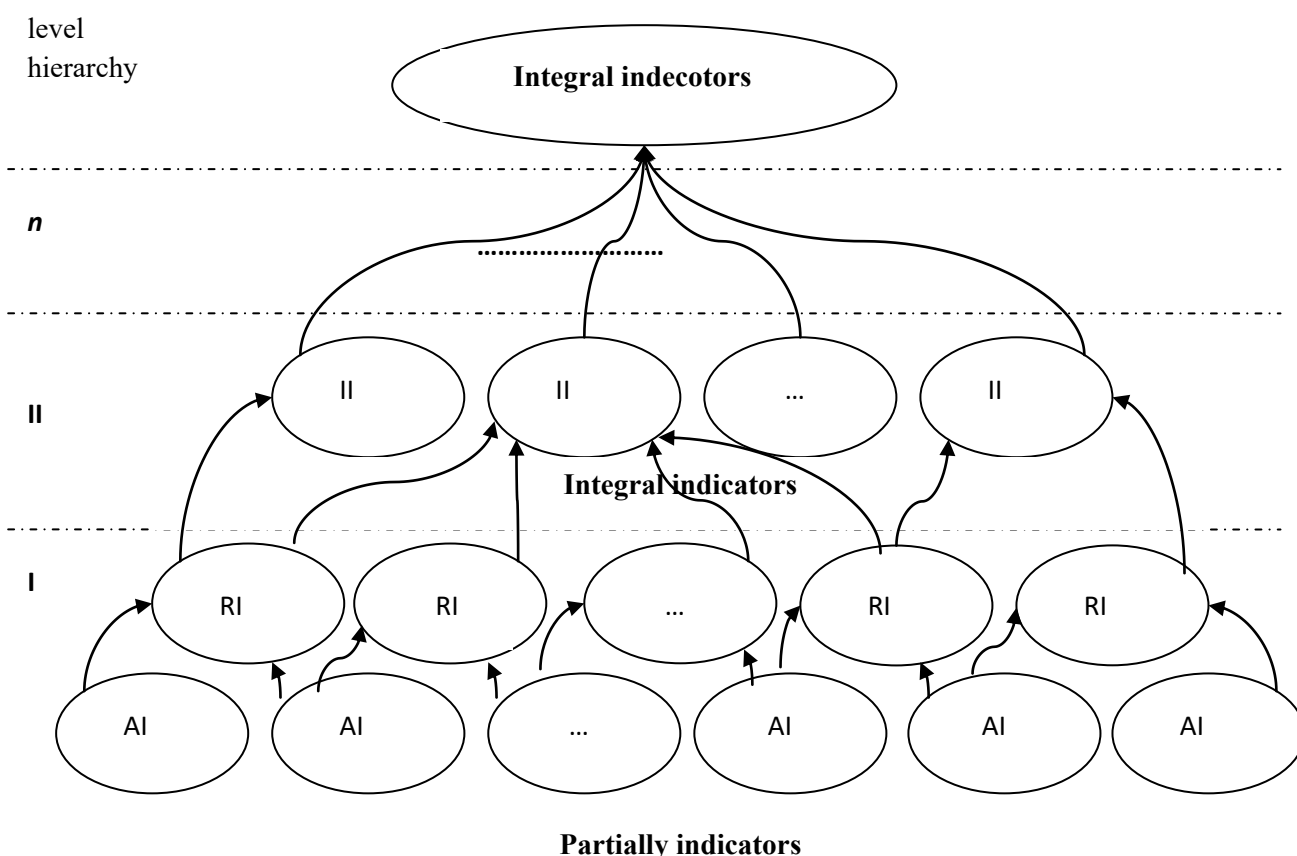
Fig.1. Hierarchy in the indicators system,
where AI - absolute indicators, RI - relative indicators, II - integral indicators

In order for the indicators system not to be bulky, the partial indicators number should be sufficient. It is recommended to use for each system component of indicators an average of 5 to 6 partial indicators. When forming a set of partial indicators, the requirements must be met, namely: 1) the system indicators adequacy to the research objectives; 2) taking into account the enterprise criteria, first of all, efficiency; 3) clarity and unambiguousness in the calculation of each indicator; 4) the presence of their reflection in accounting, financial, management reporting; 5) the existence of normative values and other recognized reference levels of the values of each indicator; 6) consistency of the indicators of influence and interconnections; 7) adaptability, that is, flexible adaptation and adjustment to changes in the external environment and in the internal environment of their own activities.

The assessment of an enterprise should be comparable. The comparison procedure involves a quantitative and qualitative comparison of various properties that manifest themselves in similarities, differences, advantages, disadvantages of two or more objects or in a dynamics on one object. Estimation of the enterprise activity on the basis of the system of indicators is based on the comparison of values in statics (in the totality of enterprises) and in dynamics at different periods of time.

In a particular enterprise, this system of partial indicators can be reduced by constructing a signifi- cant area for evaluating activities using descriptive statistics methods and multivariate statistical analysis methods such as factor analysis, canonical analysis [5].

To determine the causal relationships between the features and to identify the latent, integral complex characteristics of the industrial enterprises, it is necessary to describe the elementary features of the enterprises that were investigated by the tools of descriptive statistics. It should be noted that many scientists believe that the necessary initial stage in the nature study of the functioning and development of the object in the economy is the analysis of observations or data of its elementary features, which involves the analysis of the logical and value random variability of the sign, formed in the indicator, the detection of logical variability in the background random variability [5]. Typically, descriptive statistics tools include position statements that describe the position values of the sign value on the numerical axis, among them: selective mean, selective median, mod, minimum and maximum value of the sign, percentiles, in particular quartiles. Also, the descriptive statistics tools include scatter indicators, which describe the degree of spread of value values relative to their center and characterize the degree of their variability, namely: variance, standard deviation, velocity, interquartile velocity. The coefficients of asymmetry and excess are used to characterize the distribution of the values of the value of the sign. The degree of variability of the values of indicators characterizes the coefficient of variation.

To determine causation relationships that reflect latent factors and form mechanisms in the economy, it is advisable to use factor and canonical analysis that are implemented in the software environments MatLab, Statgraphics, Statistica.

In the factor analysis, based on the existence of a tight correlation between the indicators, identify the causes that are detected in latent factors having the same informativeness as the system of indicators.

For example, the large industrial enterprises internal environment factors such as Join-stock corporation TURBOATON are described by indicators, which are often systematized in two aspects, namely, factors of production and finance [8]. Indicators reflecting production factors include indicators such as return on assets (q_1), labor costs (q_2), resource productivity (q_3), turnover rate of inventories (q_4), depreciation factor of fixed assets (q_5), period of one stocks turnover(q_6), costs for 1 UAH of products (q_7), the finished products turnover coefficient (q_8), the operating cycle duration (q_9), the staff members average number of the accounting staff (q_{10}), the redundancy rate (q_{11}). The financial factors indicators are recommended to include the following indicators: the turnover payables rate (z_1), the equity turnover coefficient (z_2), the current liquidity ratio (z_3), the absolute liquidity ratio (z_4), the rapid liquidity rate (z_5), the equity maneuverability coefficient (z_6),the borrowing capital concentration coefficient (z_7), the financial autonomy coefficient (z_8), the coefficient of financial stability (z_9), the asset mobility coefficient (z_{10}), the assets payback period (z_{11}),the maturity of receivables (z_{12}), plowback ratio (z_{13}), reinvestment. coefficient(z_{14}). These indicators are an informative basis for developing a cognitive map for the production and financial factors influence diagnosis on the export-import activity environment efficiency of the enterprise, which is quantitatively determined by the integrated indicator of the export-import activity efficiency level of the enterprise and an integrated structural efficiency dynamics indicator of this activity.

To determine the production and financial factors influence on the overall export-import activity efficiency level of the enterprise (I_p) and the structural dynamics of this efficiency (I_s), it is first necessary to calculate the factors pair correlation coefficients with these integral indicators. Initially, factors were selected in that way, that the pair correlation coefficient with integral indexes is greater than 0,4. Thus, according to the level and significance of the pair correlation coefficients to the overall efficiency level and the export-import activity Join-stock corporation TURBOATON, the following factors are affected: return on capital, return on assets, depreciation factor of fixed assets, expenses for UAH 1, turnover rate of finished products, coefficient of staff redundancy, and also: coefficient Accounts payable, current account ratio, rapid liquidity ratio, asset mobility ratio, turnover rate of receivables, repayment period where bonds debts. The structural dynamics of the efficiency of the export-import activity of Join-stock corporation TURBOATON are influenced by the following factors: return on capital, return on assets, depreciation rate of fixed assets, costs per 1 UAH of products, as well as the ratio of current liquidity, the coefficient of absolute liquidity, the rate of rapid liquidity, the asset turnover ratio, the turnover rate of receivables , maturity of receivables. So it turned out that there are systemic factors that affect the efficiency of the company's export-import activity, namely, return on assets, resource efficiency, depreciation factor, costs of UAH 1, current liquidity ratio, rapid liquidity ratio, asset mobility ratio, turnover rate of receivables, maturity receivables.

Next, in order to detect the latent causes that form the appropriate state of the export-import activity's efficiency, the enterprise must calculate a factor analysis that will detect those latent mechanisms that occur at the enterprise and determine the appropriate state of efficiency. As a result of the calculation we have the following equations of latent factors of the general level of efficiency:

$$F_{q1}^{Ip} = -0{,}924q_1 - 0{,}919q_3 + 0{,}56q_5 - 0{,}016q_7 + 0{,}811q_8 + 0{,}763q_{11}$$;

$$F_{q2}^{Ip} = -0{,}33q_1 + 0{,}193q_3 + 0{,}587q_5 - 0{,}945q_7 + 0{,}537q_8 + 0{,}544q_{11}$$

these factors account for 86,981% of changes in the values of production factors;

$$F_Z^{Ip} = -0{,}69z_1 + 0{,}875z_3 + 0{,}949z_5 + 0{,}728z_{10} + 0{,}972z_{12} - 0{,}934z_{13},$$

This factor of 74,84% explains the change in the financial factors values . On the analysis basis of the factor load ratings of each indicator, we determine that the first latent factor of the overall export-import activity efficiency at an enterprise is the inefficient use of resources in the enterprise, the second is obsolete fixed assets, and the third, an increase in the level of enterprise asset management in order to reduce the turnover of accounts receivable.

According to the similar calculations of the latent structural dynamics factors of export-import activity efficiency, we have:

$$F_1^{Is} = 0{,}9q_1 + 0{,}926q_3 - 0{,}593q_5 - 0{,}046q_7$$;

$$F_2^{Is} = 0{,}366q_1 - 0{,}149q_3 - 0{,}673q_5 + 0{,}929q_7$$;

$$F^{Is} = 0{,}953z_3 + 0{,}916z_4 + 0{,}96z_5 + 0{,}714z_{10} + 0{,}986z_{12} - 0{,}857z_{13},$$

these factors account for 87,368% change in the values of production factors;

this factor at 81,403% explains the change in the values of financial factors. Based on the factor load rating of the partial indicators, we determine that the first latent factor of the structural dynamics of the efficiency of the export-import activity Join-stock corporation TURBOATON is to increase the use of production capacities, the second is the rationality of production costs, and thirdly, to increase the level of asset management of the enterprise in order to increase absolute liquidity.

Determination of the relationship between systems of random variables, which are indicators is possible thanks to the mathematical method of multidimensional statistical analysis - canonical analysis. This relationship is determined by new variables - canonical variables that are calculated

as linear combinations of initial attributes. New canonical variables are selected in such a way that new coordinates directly indicate the correlation value. In each group, linear combinations of initial values with the maximum correlation are searched. They will be the first new system coordinates. Then, in each group, the following linear combinations are considered, in which the correlation is greater than between any other linear combinations. The construction continues until new coordinate systems are fully developed

In the process of canonical analysis, the initial variables are brought to a standardized form, therefore, the coefficients in the equations for canonical variables characterize the effect of the corresponding initial variables (indicators) on canonical variables, which allows for the ranking of signs or indicators by force of influence.

Thus, the analysis of weight coefficients in models of canonical analysis confirms the causal relationship between the influential indicators, which were distinguished by factor analysis.

Further studies of the relationship in the enterprise performance indicators system should be continued by analyzing the pairwise relationship between the indicators with the help of the calculated Pearson correlation coefficient. It is recommended to take even correlation coefficients higher than 0,5. The analysis of these correlation coefficients shows which indicators in the system are not interrelated in the system either within its component or with the indicators of other components.

Consequently, when forming a symbolic space for evaluating the activity of an enterprise, the logic of the stages presented in fig. 2

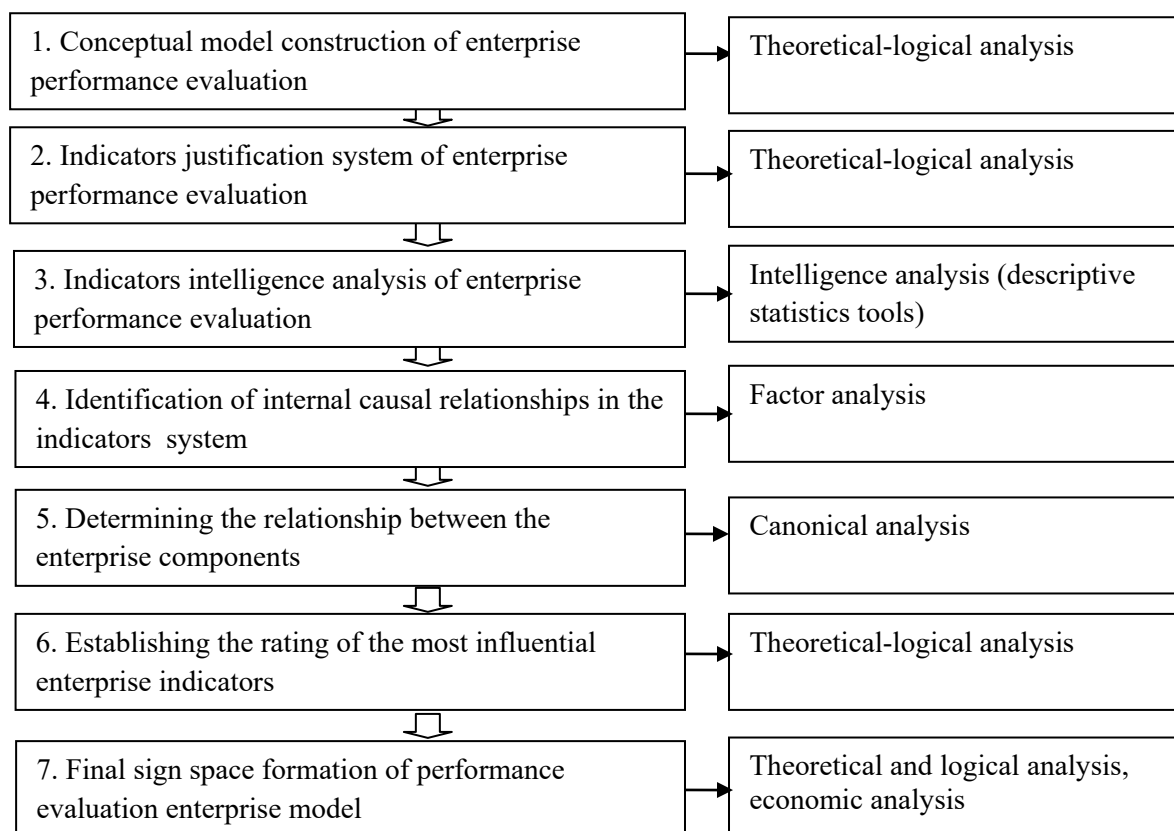| 1. Conceptual model construction of enterprise performance evaluation	Theoretical-logical analysis
2. Indicators justification system of enterprise performance evaluation	Theoretical-logical analysis
3. Indicators intelligence analysis of enterprise performance evaluation	Intelligence analysis (descriptive statistics tools)
4. Identification of internal causal relationships in the indicators system	Factor analysis
5. Determining the relationship between the enterprise components	Canonical analysis
6. Establishing the rating of the most influential enterprise indicators	Theoretical-logical analysis
7. Final sign space formation of performance evaluation enterprise model	Theoretical and logical analysis, economic analysis

Fig. 2. Logic of the stages formation of a sign modeling space for assessing the enterprise activity

Conclusions

The simulation process is complete if it based on the calculated mathematical model, it is possible to justify and adopt an effective management solution. Therefore, the enterprise performance evaluation modeling involves the final stage of the managerial decision development to improve the efficiency of the enterprise and improve its management. Here, as a reference point, a notable space of the model of enterprise performance evaluation should be used.

References

[1] Robert S. Kaplan, David Norton David. Harvard Business Review. 1996 – 2018).
https://hbr.org
[2] Marshall W. Meyer. Rethinking Performance Measurement: Beyond the Balanced Scorecard. Cambridge University Press. 2003.
[3] P. Niven. Diahnostyka zbalansovanoyi systemy pokaznykiv: Pidtrymuyuchy maksymalnu efektyvnist [Diagnosis of Balanced Scorecard: Maintaining maximum efficiency], Dnipropetrovsk. Balance Business Bouks, 2006.
[4] Nils-Göran Olve, Jan Roy, Magnus Wetter. Performance Drivers: A Practical Guide to Using the Balanced Scorecard. Wiley, 2001.
[5] V. C. Ponomarenko, L.M. Malyarets, L.M. Analiz danykh u doslidzhennyakh sotsialno-ekonomichnykh system [Analysis of data in researches of socio-economic systems], monograph, Kh.: VD "INZHEK", 2009.
[6] Hubert K. Rampersad. Total Quality Management: An Executive Guide to Continuous Improvement. Springer, 2001.
[7] L. Malyarets, M. Draskovic, N. Proskurnina, O. Dorokhov, V. Vovk. Analytical support for forming the strategy of export-import activity development of enterprises in Ukraine. *Problems and Perspectives in Management*, Volume 16, Issue 3, 2018.
[8] Join-stock corporation TURBOATON (official site). http://www.turboatom.com.ua/)

Lyudmila SABADASH
Simon Kuznets Kharkiv National University of Economics
Department of Higher Mathematics, Economic and Mathematical Methods
Nauki avenue, 9-A, Kharkiv, 61166
UKRAINE
malyarets@ukr.net

Pavlo GRYNKO
Simon Kuznets Kharkiv National University of Economics
Department of Higher Mathematics, Economic and Mathematical Methods
Nauki avenue, 9-A, Kharkiv, 61166
UKRAINE
malyarets@ukr.net

Cryptographic Solution with Poly-Alphabetic Cipher

Victoria Sărăteanu, Paula Ioana Barna, Andrei Torok, Maria-Alexandra Bogdan

Abstract

This article aims to establish a prime contact for first year students in computer science with cryptography research field and to familiarize themselves with terms and methods from the field. For this purpose in this project we approached some widely spread cryptographic techniques, namely: Caesar's Cypher, Vigenere's Cypher and Pseudorandom Number Generator. We have successfully implemented the tasks by using C++, C#, Net Framework and Windows Form App.

1 Introduction

This article aims to establish a prime contact for first year students in computer science with cryptography research field and to familiarize themselves with terms and methods from the field. For this purpose in this project we approached three widely spread cryptographic techniques: Caesar's Cypher, Vigenere's Cypher and Pseudorandom Number Generator. We have successfully implemented the tasks by using C++, C#, Net Framework and Windows Form App.

1.1 Caesar's Cypher

Gaius Julius Caesar (cca. 13.07.100 b. Ch - 44. b. Ch.) was a roman politician, general, and military strategist that proved his genius abilities both in politics and battles, and among others, created the encryption method that bares his name: Caesar's Cypher [2].
Normal text in the article: Times New Roman, 11pt, Character spacing Extended by 0.2 pt
Caesar's cypher was initially created to protect military high - priority and importance messages. Along time, the cypher was used in a high number of military actions, and, one of the most well-known cases was in 1915, when Russian army used it in order to replace more difficult cyphers that proved to be too complicated for them at that time. Meanwhile, German and Austrian armies didn't find difficult to decrypt the encrypted messages with Caesar's Cypher.
In Cryptography, Caesar's Cypher is also known in other ways, like: shift code, Caesar's code, or Caesar shift, and is one of the most easy decryption techniques [3].
Cryptology aims to defend secret data and confidential information using mathematic methods and Information Technology means. Cryptology has two sides, defensive and offensive.
Cryptography is the defensive side of cryptology that aim to conceive encryption systems and their terms of use, while Cryptanalysis is the offensive side of Cryptology that study own cryptographic systems in order to give those specific attributes so that systems could fulfil their mission successfully.

Also, Cryptanalysis can study third party cryptographic systems in order to break the code and gain the necessary information held by cryptographic systems.

Even though Caesar was the first one to use an encryption systems, it is known others, less popular substitution-based encryption systems.

The algorithm of Caesar shit is based on a mono-alphabetic encryption system for which the text should exclusively contain Latin letters and the encryption key is a whole number.

To each letter of the key text is attributed a lexicographic order x. For encryption letters are replaced by *x + a whole number*, and for decryption there is used the reverse command: x-whole number.

In other words, to encrypt a message all you have to do is to change the lexicographic order up to an absurd meaning, and for decryption, the reader had to understand the replaced letters in the lexicographic order.

E.G.

Initial Message: **ANA ARE MERE DE LA BUNICA SA**

Encrypted Message: **DSD DUÎ RÎUÎ IÎ QD GZSMHD VD**

This example defines a simple move of three letters forward from the original lexicographic order, and the encryption is represented by formula x+3.

Shift code can be broken even by simply having the crypto-text, and thus, can be taken in consideration two ways:

- The attacker knows that the encryption method was made by a simple substitution, and not necessarily a Caesar system. So that in the decryption process it is very probably to logically deduct the algorithm and notice the rhythm it was used in the crypto-text
- The attacker is aware that it was used a Caesar code in the crypto-code, and is able to break the code much more simple, as the algorithm is limited by approaching the brute force. In case of brute force approach, means the identification of letter distribution in the alphabet the cryto-text was written.

Even so, there are proves that Caesar used more complicated formulas to encrypt messages, and Aulus Gaelius makes reference in a treaty about Caesar code:

" There is even a treaty written in a very ingenious way by Probus that depicts the secret meaning of Caesar's epistoles" - Aulus Gaelius, 17.9.1 – 5

1.2 Vigenere's Cypher

Blaise de Vigenère (5 April 1523 – 19 February 1596) was a French diplomat, cryptographer, translator and alchemist. The cypher was named after Blaise de Vigenère, but it was discovered earlier, in the 16th century, by Giovan Battista Bellaso. However, Blaise de Vigenère improved Bellaso's work and developed a stronger auto-key cypher.

The cypher was indecipherable for three centuries until 1863. This method of encrypting is using a series of interwoven Caesar's cyphers.

In a Caesar cypher, each letter of the alphabet is shifted along some number of places. For example, in a Caesar cypher of shift 3, A would become D, B would become E, Y would become B and so on. The Vigenère cypher has several Caesar's cyphers in sequence with different shift values [4].

To encrypt, a table of alphabets can be used, named Vigenère square or Vigenère table. It has the alphabet written out 26 times in different rows, each alphabet shifted cyclically to the left compared to the previous alphabet, corresponding to the 26 possible Caesar ciphers. At different points in the encryption process, the cipher uses a different alphabet from one of the rows. The alphabet used at each point depends on a repeating keyword.

The person sending the message chooses a keyword and repeats it until it matches the length of the plaintext. Each row starts with a key letter. The rest of the row holds the letters A to Z (in shifted order). The encrypted letter is the letter at the intersection of [key-row, message-column].

For example:
Plaintext: **WHITE**.
Keyword: **LEMON**.
Cypher-text: **ILUHR**.

Vigenere can be also described algebraically. If the letter A→Z are taken to the numbers 0→25, and addition is performed modul 26, Vigenere encryption E using key K can be written as:
$C_i=E_k(M_i)=(E_k+M_i)$ *modulo 26.* Figure 1 shows a Vigenere table.

Fig. 1 – Vigenere Table

2   Proposed solution

2.1 Problem definition

In this project we aim to implement an algorithm that converts a text read from the keyboard into unreadable format. In this scope we define the following statement: Develop a program that asks the user to enter a text from the keyboard (TO), a seed (s) to generate the encryption key (k), and display the cipher text (TC) after the following algorithm:

- Transforms the original text into ASCII code (T_AS). ASCII (American Standard Code for Information Interchange) is a character encoding system. Each character read from the keyboard is encoded using the decimal ASCII code to have a common language on all the devices that use the information;
- It converts T_AS (obtained previously) from base ten into base two (from natural integers in zero and one) and the result obtained is denoted by T_B. Note that seven memory bits will be used to store a value (seven values of zero or one).

- To generate the encryption key (k), the seed value (s) entered by the user is considered. This must be a native natural number. A pseudo-aleator algorithm generates a coding key using c cstdlib library. The ascii code of the obtained key is converted from base ten to base two. The result is note with k_B.
- Cases where k_B is less than or greater than T_B should be discussed. If k_B <T_B, k_B will multiply.
- The XOR operation is run between k_B and T_B and it will result the encrypted text in binary (in the base two) marked with T_CB. The XOR operation is described next.
- T_CB is converted from binary (zero and one, base two) in integers (on base ten). The numbers obtained are the ASCII code of the characters (printable or unprintable) of TC.
- Converting ASCII code into characters and displaying the text.

2.2 Implementation

In the first stage of the project we developed the algorithm using c++ programing language because we were more familiarised with its usage. The next step we wanted to give a friendly user interface for the proposed solution so we had to convert the code into C# and put it through Visual Studio (Windows Form App, C#, Net Framework).

For functionality we used:

- a main (parent) form for presentation;
- a second (child) form for executing the code;
- two buttons for selecting the desired method;
- one rich text box to fetch and store for further use the inserted text (for encrypting);
- one rich text box to fetch and store for further use the key used for encrypting the text.

Some code examples that we used in this project are presented as fallows.

```
public partial class FormChild : Form{
    public static int toBin(int num){
        int bin = 0;
        int rem;
        int i = 1;
        while (num != 0){
            rem = num % 2;
            num /= 2;
            bin += rem * i;
            i *= 10;
        }
        return bin;
    }
}

public static int toDec(int num){
    int decimalNum;
    int i;
    int rem;
    decimalNum = 0;
    i = 0;
    while (num != 0){
        rem = num % 10;
        num /= 10;
        decimalNum += rem * Convert.ToInt32(Math.Pow(2, i));
        ++i;
    } return decimalNum;
}
```

The fallowing code shows the implementation of Vigenère cipher.

```
string msg;
msg = TextBox.Text;
string key;
key = Key.Text;
TextBox.Text = null;
Key.Text = null;
int keyLen = 0, i, j=0;
keyLen=key.Length;
int encryptedMsg[255];
//encryption

for (i = 0; i < msg.Length; i++){
    if('A'<=msg[i] && msg[i]<='Z'){
        encryptedMsg[i] = ((msg[i] + key[j]) % 26) + 'A';
        j++;
        if(j==keyLen)j=0;
    }
    else encryptedMsg[i]=msg[i];
}

encryptedMsg[i]='\0';
MessageBox.Show(encryptedMsg);
```

3   Conclusions

Encryption is a very efficient tool that has a major impact on tech development and personal development. In this paper we get familiarised with the basic algorithms of cryptography and their basic form of implementation. For our future work we intend to approach more complex solutions of cryptography and we aim to develop an optimized version of modern cryptography techniques.

References

[1] David Kahn, *The Codebreakers — The Story of Secret Writing*, 1967. ISBN 0-684-83130-9.
[2] F.L. Bauer, *Decrypted Secrets*, ediția a doua, 2000, Springer. ISBN 3-540-66871-3.
[3] Chris Savarese and Brian Hart, *Caesar's code*, 1999.
[4] Vigenère, *Blaise de—Treatise on ciphers, or secret ways of writing*, 1586.

Victoria SĂRĂTEANU
Aurel Vlaicu University
Departament of
Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: victoria.sarateanu1@gmail.com

Paula Ioana BARNA
Aurel Vlaicu University
Departament of
Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: paula.barna@yahoo.com

Andrei TOROK
Aurel Vlaicu University
Departament of
Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: andrei_tk@yahoo.com

Maria-Alexandra BOGDAN
Aurel Vlaicu University
Departament of
Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail : bogdan.alexandra99@yahoo.com

Digital Image Analysis Based on Texture Features

Ira Tuba, Una Tuba

Abstract

In recent years analogue images have been completely out of the use due to the great expansion of digital images. Cheap and affordable equipment for creating and processing digital images has contributed significantly to their widespread. One of the common tasks in applications that include digital images is image segmentation. Segmentation can be done in numerous ways and one of them is based on the texture features. In this paper we present a software framework for texture analysis. As texture descriptor we used different local binary pattern variations. We analyzed differences and compare the quality of them. We used different benchmark images for testing the texture descriptors.

1 Introduction

In recent years analogue images have been completely out of use due to the great expansion of digital images. Cheap and affordable equipment for creating and processing digital images has contributed significantly to their widespread. Digital images are in use in numerous and diverse areas such as medicine, journalism, astronomy, biology etc. Many of the advantages of digital images have led to great advancements in some very important fields. Probably the greatest advantage of digital images is that they are easy to process using mathematical formulas and models. Some details that are invisible to human eye now can be edited and examined. There are many ways and reasons to analyse digital image but in this paper focus is on digital image analysis based on texture features. Image texture analysis is a great problem in computer vision and pattern recognition. Segmentation is one way of describing texture of the image. The benefits of characterizing the various textures post segmentation is that it allows to:

1) make better context of what the scene is,
2) compress the image,
3) in medical and scientific applications, it allows to determine the material in each segment.

Over the years has been developed simple but efficient methodology for texture analysis based on Local Binary Patterns (LBP). Due to the efficiency and simplicity of the LBP operator, numerous variants have been introduced and used in different applications.

In this paper we implemented a framework for testing LBP operator with the square neighbourhood. The framework creates LBP image and corresponding histogram.

2 Digital image processing

Digital images are represented as matrix of numbers. The digital image is a two-dimensional function f (x, y), where in the point with coordinates x and y the value of the function represents the intensity at that location, while the domain of the function is definite [1]. Analysing of digital images includes algorithms for transforming the digital image record. The processing of digital images is very simple. It is implemented using a mathematical formula on a matrix representing a digital image record. Every change is reversible. If formula is applied to an image, the image returns to the original state by applying the inverse formula. For evaluating the quality of the processed image key factor is the man. Images are processed for different purposes, from a better aesthetic experience to easier recognition of some features on it.

Digital image processing can be divided into three levels. The first level, also called a low level of processing, includes some kind of image enhancement without knowing anything about what is on the picture. Processing at this level includes noise removal, brightening, darkening, contrast enhancement, etc. All transformations are done on global level. Second or middle level processing deals with some kind of segmentation, i.e. finding some of the characteristics of the image. This level of processing includes operations such as skin color recognition, face recognition, finding non-sharp parts of the image, etc. The needed parts of the image are found based on the arrangement and relationship between pixels and their intensity. The third, high level of processing involves some kind of artificial intelligence. For example, the previous level had only to recognize the face in the picture but the third level should also recognize whose face is it.

2.1 Segmentation

There are different areas of digital image processing. Image segmentation is done when it is necessary to extract some of its characteristics. This can be achieved by image binarization. The object that needs to be highlighted will be white, while the background is black. This is often required in the processing of medical digital images.

Segmentation is splitting the whole image into semantically interpretable regions. This is one of the crucial steps in case of digital image analysis. The outcome of image segmentation very much relies upon precision of feature measurement. Segmentation is splitting the whole image into its ingredients and taking out object of interest. There are many techniques for image segmentation but some of them are more popular and regularly used [2]. These methods are classified as follows:

- Region Based Segmentation
- Edge Based Segmentation
- Threshold Segmentation
- Cluster Based Segmentation

Region based segmentation systems is grouping together pixels with identical features into segments [4]. This technique comparing to others is very simple but efficient. Region based segmentation is dividing image based on some specification such as object, color or intensity. The segmentation methods based on region are divided into the following categories: region splitting, region growing, and region merging [3].

Edge based segmentation is segmentation technique that is splitting the image by finding differences in image pixels or intensity [5]. This is basic step in process of image segmentation. The result is digital image in binary form. Two primary techniques which are used for segmenting image via edge detection are gray histogram and gradient [6].

Threshold segmentation is primary technique for image segmentation. That method is used for discerning background from foreground. The gray scale image can be converted into binary image by choosing an acceptable threshold value. The algorithms based on thresholding can be obtained physically by having some priori knowledge or repeatedly by formation of image [7], [8].

3   Texture analysis

Texture analysis methods have been divided into two groups [9]. The first one is called statistical or stochastic approach. This method treats texture as a statistical appearance. The formation of a texture is described with the statistical properties of the intensities and positions of pixels. This formulation of texture is based on a model in which a texture is a sample of a two-dimensional stochastic process describable by its statistical parameters. The interactions between pixels are mostly of a local type. Texture is characterized not only by the gray levels of pixels, but also by local gray value "patterns".

Second category is called structural approach. This method introduces the concept of texture primitives named texels or textons. To describe a texture, a list of texels is necessary as well as relationships between them. The aim is to describe complex structures with simpler primitives like graphs. Structural texture models work well with macrotextures with clear constructions.

The LBP method can be impersonates unifying approach. Instead of trying to explain texture formation on a pixel level, local patterns are formed. Each pixel is labeled with the code of the texture primitive that best matches the local neighborhood. So each LBP code can be seen as a micro-texton. Local primitives detected by the LBP include spots, flat areas, edges, edge ends, curves etc.

4   LBP and its modifications

Local binary pattern (LBP) was introduced for measuring local image contrast and later adopted as statistical and structural texture descriptor [11]. The LBP define textures by using textons which is histograms of texture elements. The texture is described for each pixel by using the local structure. Binary code is extracted based on the intensity level differences between neighborhood pixels. Pixel's intensity level is used as a threshold value for pixels around it. Binary code is defined by reading obtained values in certain directions and they represent different curved edges, spots, flat areas and similar features in an image. When texture codes are calculated for each pixel in a given image, the histogram should be created that represents the frequency of the occurrences of each pattern, i.e. binary code. This histogram is a vector that is the result of applying LBP. The dimension of the vector is determined by the neighborhood size that was used for building the patterns.

LBP has been proved to be rather good texture descriptor while computationally not expensive. Due to its quality, it attracts the attention of scientists and over the years, various modifications have been proposed [10]. The original LBP operator builds binary codes or patterns by using 8-neighborhood, i.e. eight pixels that surround the central pixel. Original LBP is rotation variant which is not always a useful feature for texture descriptor. On the other hand, it is invariant to monotonic changes in intensity level.

The LBP operator is defined by symmetric neighborhood set of P pixels of the central pixel g_c. The neighborhood is defined by radius R which represents the distance between central pixels and pixels g_i where i=0, 1, 2, …, P-1 that build the pattern. The original LBP is with R=1. Mathematically, LPB operator is calculated by the following equation:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p,$$

(1)

where

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0. \end{cases}$$

(2)

The number of different patterns is 2^P which will be the size of a histogram, i.e. dimensionality of the vector that is the output of LBP operator.

Rotation variant texture descriptor is not always desirable and in order to overcome that problem, rotation invariant versions of the LBP have been proposed. A simple modification is to perform a bitwise shift operator before assigning the code. If the P neighborhood is used that each pattern is shifted P-1 times and the smallest pattern obtained during the shifting is used as a final code.

Another rotation invariant LBP operator that is commonly used is the uniform local binary pattern. One binary pattern is uniform if it does not contain more than two spatial transitions. Spatial transition is bitwise changes from zero to one or other way around, from one to zero. Besides that, this is a rotation invariant model of LBP, it also reduces the size of the output vector. In the case of original LBP, when 8-neighborhood is used, the size of the output vector is 2^8=256. By using uniform LBP and the same neighborhood structure, the number of different patterns is reduced to 59 since there are 58 binary codes with zero, one or two spatial transitions and 1 pattern is used for all others. This is rather important characteristics since further classification has less but equally descriptive inputs.

Other variants include using different topologies of the neighborhood. Rather commonly used topology is circle and topologies such as an ellipse, parabola, hyperbola are often used for specific applications. In this paper, we created a framework for testing LBP operator with square topology.

5 The proposed framework

Framework for digital image registration was implemented in Visual Studio 2017, C#, NET Framework 4.7.1. The original and uniform LBP with different size of square neighborhoods were implemented in the initial version.
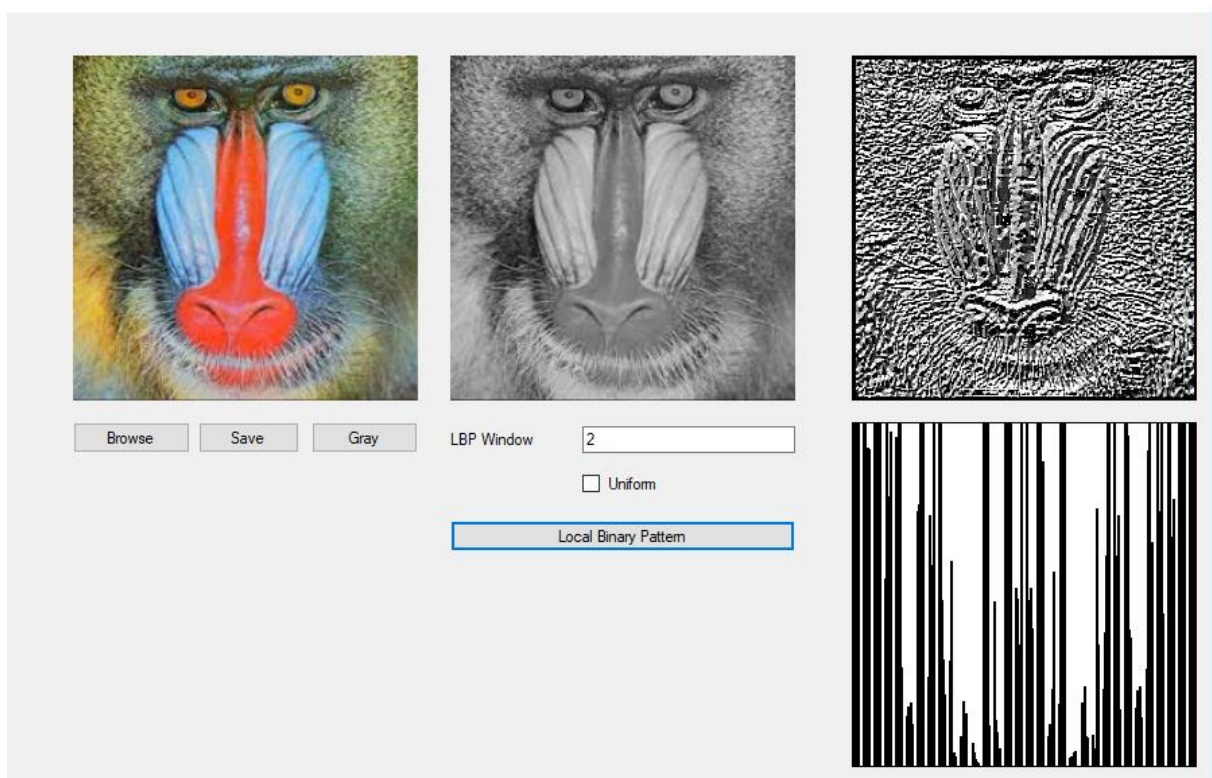

Fig. 1 GUI of the proposed framework for testing LBP

The GUI of the proposed framework for testing LBP operators is presented in Fig. 1. The first image represents the original image, the second one is a gray scale version of the chosen image and the third one represents the LBP image. Under the LBP image, a histogram of LBP is presented. For choosing

an image, there is a *Browse* button. Button *Gray* transform chosen image into an 8-bit grayscale image. The button *Save* enables to save resulting LBP image. Before transforming the image to LBP, it is necessary to choose distance R that will be used. Original LBP is with the distance R=1.

We tested our framework with different images. In Fig. 2 is shown original image and LBP images with R=1,2 and 3 are presented in Fig. 3.

Fig. 2 Test image 1
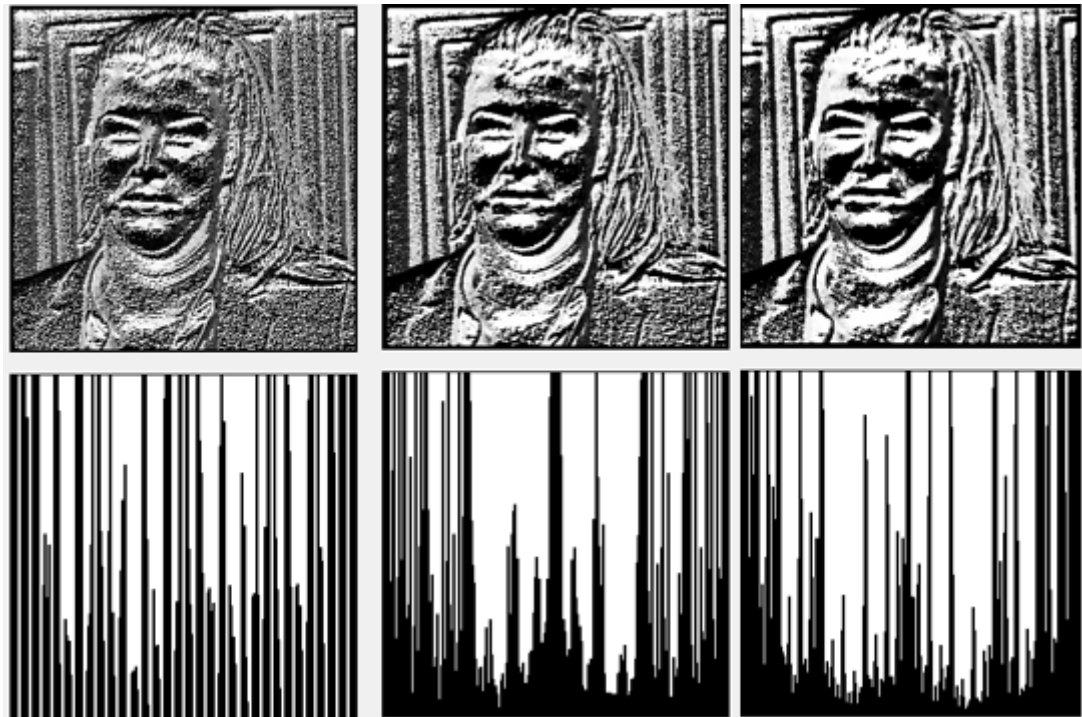
Fig. 3 LBP images for test image 1 and distances 1, 2 and 3

As it can be seen when different distances are chosen for square LBP operator, different features are emphasized in the image. The main features in the image are still visible but based on the application different distances can better present the searched characteristics.

Since LBP operator represents texture descriptor, we used two benchmark texture images and tested our framework. The used images and corresponding LBP images and histograms are presented in Fig. 4.
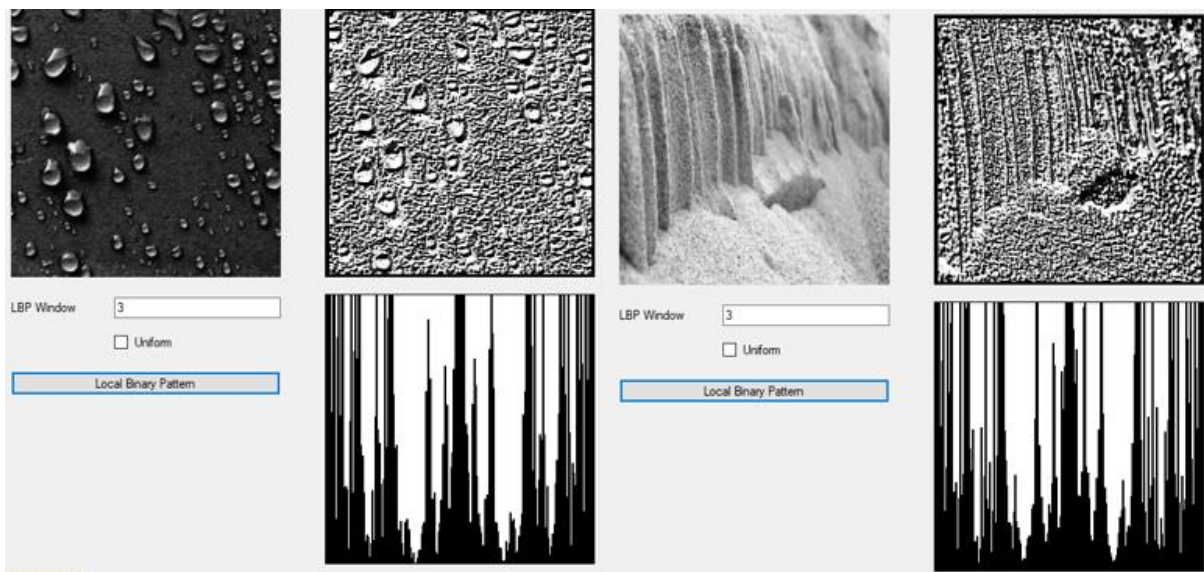
Fig. 4 Different texture images and their corresponding LBP images

6 Conclusion

Texture analysis represents an important task in numerous application. In the past various different texture descriptors were proposed and they were applied to many different practical problems. One of the widely used texture descriptors is local binary pattern along with its variationa, In this paper, we presented a framework for experimenting with the local binary pattern. The original and uniform LBP were implemented in the initial version described in this paper. In further work, it was planned to implement more neighborhoods, also to provide some histogram analysis. In order to expand to usage of the framework, texture matching methods can be implemented.

References

[1] Gonzalez, Rafael C., and R. E. Woods. "Digital image processing", 3rd Edition, Pearson Education, 2011.

[2] Aly, A. A., Deris, S. B., and Zaki, N. "Research review for digital image segmentation techniques", *International Journal of Computer Science & Information Technology*, 3(5), pp. 99-106, 2011.

[3] Kaganami, H. G., and Beiji, Z. „ Region-based segmentation versus edge detection", In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 1217-1221, 2009.

[4] Wang, Z., Jensen, J. R., amd Im, J. „An automatic region-based image segmentation algorithm for remote sensing applications", *Environmental Modelling & Software*, 25(10), pp. 1149-1165, 2010.

[5] Uemura, T., Koutaki, G., & Uchimura, K. „ Image segmentation based on edge detection using boundary code", *International Journal of Innovative computing, Information and control*, 7(10), pp. 6073-6083, 2011.

[6] Lakshmi, S., and Sankaranarayanan, D. V. „A study of edge detection techniques for segmentation computing approaches", *IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT*, pp. 35-40, 2010.

[7] Vala, H. J.,and Baxi, A. „ A review on Otsu image segmentation algorithm", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(2), pp. 387-389, 2013.

[8] Cuevas, E., Zaldívar, D., and Perez-Cisneros, M. „Otsu and Kapur segmentation based on harmony search optimization", In *Applications of evolutionary computation in image processing and pattern recognition* (pp. 169-202). 2016.

[9] Selvarajah, S., and Kodituwakku, S. R. „Analysis and comparison of texture features for content based image retrieval", *Energy*, *1*, pp. 108-113, 2011.

[10] Nanni, L., Lumini, A., and Brahnam, S. „Local binary patterns variants as texture descriptors for medical image analysis", *Artificial intelligence in medicine*, *49*(2), pp. 117-125, 2010.

[11] Lahdenoja, Laiho M., and A. Paasio. "Local binary pattern feature vector extraction with CNN," *9th International Workshop on Cellular Neural Networks and Their Applications*. IEEE, pp. 202-205, 2005.

Ira TUBA
Singidunum University
Faculty of Informatics and Computation
Danijelova 32, Belgrade
SERBIA
E-mail: ira.tuba@gmail.com

Una TUBA
Singidunum University
Faculty of Informatics and Computation
Danijelova 32, Belgrade
SERBIA
E-mail: tuba.una@gmail.com

Analysis of prevention system of Cross Site Scripting

Milan III Tuba, Stefan Ilic

Abstract

Internet has become an unavoidable part of everyday life. We are relying more and more on the Internet and Web applications and the work that is done by using Web applications (Net Banking, Shopping, etc.) is constantly growing. Thus, the interest and motivation of the attackers for the search and exploitation of vulnerability are also rapidly growing. One of the most used vulnerabilities of the web applications is Cross Side Scripting or XSS. Cross Side Scripting allows attacker to execute a script on the user's browser and to steal the valuable information such as passwords, credit card number, personal information, etc. In the case of the hijacking malicious scripts into the trusted web source, the user is most often completely unaware that he is a victim of an attack. This type of attacks is very hard to detect and yet rather easy to use. That is why it is very important to give a special attention during programming the web application to be ensuring that there are no such vulnerabilities. It became more difficult to protect web applications from this vulnerability if application have to allow user to use HTML tags (forums, emails, etc.). Unfortunately, many web sites still have this vulnerability. In this paper we will analyse several different prevention systems, test their efficiencies, and point out their qualities and flaws. Also, we will demonstrate Cross Site Scripting on the practical example and discuss how to prevent it.

1   Introduction

The rapid advancement of technology and thanks to their availability and advantages, a lot of companies move their business in digital world. According to a survey conducted by the government of Great Britain 56% of companies keep client personal data in digital form. On the other hand, 76% of companies are using Net Banking services. Because of this trend, security and protection of information systems is high priority for today business. Almost all companies use computers and Internet in their every day jobs.

One of the most popular and the most used vulnerability in web applications is Cross Site Scripting (XSS). This vulnerability is on seventh place on OWASP Top 10 2017 from Open Web Application Security Project™ foundation. Unfortunately, still a lot of websites have this vulnerability. Is hard to detect, but easy to execute, so is very important to pay particular attention to it.

Cross site scripting is vulnerability that allows attacker to inject code into a website and execute malicious script on user web browser. Script is usually written in script language like JavaScript, VBScript, Flash, etc. But most commonly is using combination of HTML and JavaScript languages. There are a few methods how Cross Site Scripting is executed depending on type of

script injection. Cross site scripting vulnerability is divided on three different methods of executing [3]:

1. Persistent or Stored XSS
2. Reflected XSS
3. DOM-based XSS

Stored or Persistent Cross Site Scripting is method of code injection into a user content, like comments, submission forms, contact forms, etc. Because of poorly implemented security user is allowed to type HTML code like comment and post it in website database. For an example script that will send cookie to attacker server:

```
<script>
window.location='http://attacker/?cookie='+document.cookie
</script>
```

When user post this type of input, web application automatically saves user input in database and after next user try to see comments website will execute following script to display comments:

```
print "<html>"
print "Comments:"
print database.Comments
print "</html>"
```

After website request form database comments, it get text from first example and looks at it as a part of his own code and execute script inside script tags. In this process user is not aware that malicious script, that is not made by website owner, is executed on his browser. This type of Cross Site Scripting is targeting every user who visit specific website. Malicious script can steal cookies from that website and take from them sensitive information, send keystrokes to his server, where can be found sensitive information like credit card number or passwords [4].

Reflected Cross Site Scripting is method where malicious script is part of victim request to website. Then, the malicious script is part of website response. This method works on websites where user input become part of URL and with that URL you can access same page again, like search engines. Attack works when attacker make malicious URL and send it to victim. For example:

```
http://serachengine/search?keyword=<script>window.location='http://attacker/?cookie='+document.cooki
e</script>
```

When victim open that link and send request, in response will be a script and victim browser will execute it. After all that attacker will have cookie information on his server. For this method it is necessary that the victims click on the link. That can be done on many ways. For example, using some URL shortener attacker can mask real URL and send it to specific victim using e-mail, instant messaging, social network, etc. Also, if attacker is targeting some larger group, he can post it on some public place, like forums, blogs, social network, etc. and wait for people to open it.

DOM-based Cross Site Scripting is a combination of stored and reflected. During this type of attack, victim browser does not process the malicious script until the website legitimate code is executed. So, attacker again have to make malicious URL and trick victim to open it. But in this case response does not contains malicious script. Firstly, the browser executes legitimate code from website response, and then malicious script is inserted into a page cause by that. And then, victim browser executes malicious script. DOM-based XSS is different from stored and reflected because there is no malicious script inserted into a website page. The legitimate script use user input to insert HTML to page. Malicious script is inserted using innerHTML that causes it to be executed.

DOM-based XSS matters because in previous methods JavaScript was not indispensable. Whole HTML is generated on server-side, so if server-side code is without vulnerabilities, then the whole website is safe for user. However, more advanced web applications do not generate much of HTML on sever then on the client-side generated by JavaScript. Update of the page is done by JavaScript, because the content has to change without refreshing page. For an example, when a page, after AJAX request, is updated.

XSS vulnerabilities can exists in websites client-side code, not only in server-side code. Thus, even if server-side code absolutely secured, the website still might be accepting user input in a DOM update after page is loaded. In this situation client-side code has allow a Cross Site Scripting attack without the fault of the server-side code.

In some specific case of DOM-based XSS attack, malicious script is never sent to the server. This is a case when malicious script is in URL fragment identifier, after the # character. Browser do not send URL fragment identifier to server, so the website has no means to see or detect XSS attack. On the other hand, client-side have access to this part and can cause this vulnerability by using it insecurely.

This type of incident is not tied to URL fragment identifiers specific. There are an others user input that is invisible to the server, like HTML5 features, IndexedDB and LocalStorage.

2   Prevention

Since the XSS is code injection, wrong interpretation of user input as a part of website code, it is necessary to securely handle user inputs. Commonly used methods for prevention of Cross Site Scripting are [1]:

1. Data validation
2. Filtering
3. Escaping

Data validation is filtering user input to remove all malicious parts of it. Most commonly used type of data validation is permitting some HTML tags like <b>, <i>, <u>, etc. and forbidding others like <script>.

There are two main strategies in data validation. These strategies are [5]:

1. Classification
2. Outcome validation

Classification strategy is done by using either blacklist or whitelist. Blacklisting is a method where web developer makes a list of forbidden inputs. In our work we made blacklist with script tag. Essentially, PHP script that insert data into a database check if there is a script tag and, if it is found, remove it. We tried to insert following script:

```
<p id="demo"></p>  <script> document.getElementById("demo").innerHTML = "Hello JavaScript!";
</script>
```

Since we made a blacklist with this script tag, user's input is inserted into database without it as it is demonstrated in Fig. 1.

Fig. 1 Insert into the database with blacklist

It looks like a good method, but if the attacker changes anything, for an example types a capital S (<Script>) as it is demonstrated in Fig. 2, or if any other small change is made, the blacklist will fail.

Fig. 2 Example of blacklist failure

The conclusion is that the blacklist has to be very complex in order to cover all possible inputs. Even if weeks have been spent for making a perfect blacklist, it can fail with any new feature added to the browser. For an example, HTML5 added new feature *onmousewheel* so all blacklist made before HTML5 are now useless. Basically, it is strongly discouraged to use this method. On the other hand, whitelist represents the opposite of the blacklist. Instead of defining all malicious input, you can just define accepted input, for an example, remove all tags except <b>. This method is much simpler, because there is a limited number of tags that you want to allow. Also, if there is a new feature, whitelist will still be good as before.

The next method that can be used to prevent XSS is outcome validation reject inappropriate input which is a simple method for implementation. It is usually used to check if the input meets certain specifications of that field. For an example, if the input in the email field is matching to email address, or in other words, if input has the following form: *example@website.domain*. The only use of this method for preventing XSS is to check if the appropriate data is inserted. Web developers should use this method for every input where it is a possible, not just because of XSS but also because it prevents some bugs and errors in web application.

Filtering is a method where is each tag from user input has been removed. So, basically everything that can be considered as part of a website code should be removed. That is a very good method for securing web application, but still it is imperfect. For an example, on some web pages perhaps you have to use some of filtered code. Imagine search engine where you cannot search HTML tags. It will be useless for web developers. Or a forum which does not allow you to change font or post a picture. Also, there is another way of XSS scripting that does not include HTML tags to be executed. On Google learning platform for XSS (https://xss-game.appspot.com), on level 3 input that will execute script does not have any tag (' onerror='alert(1)'; Fig. 3). Thus, this method will not work in that case.
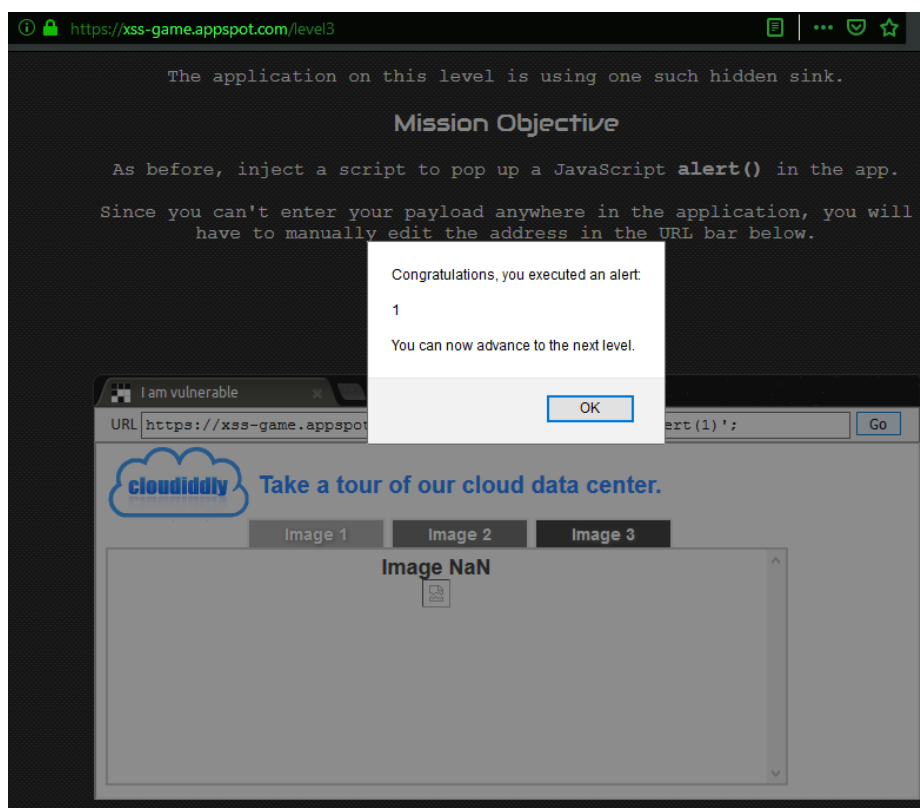

Fig. 3 Failure of the filtering method

The last method that we implemented and tested in this paper is escaping. This method is the most secure one of all the named methods. In this method every special character is transformed to the HTML encode (e.g. '<' and '>' are transformed into '<' and '>'). When you remove all special characters, user's input *<script>...</script>* will now look like *<script>...</script>*. Thus, the browser will not look at any part of the user input as HTML but it will just print it same as user inserted it (Fig. 4).

Secured	`<p id="demo"></p> <script> document.getElementByld("demo").innerHTML = "Hello JavaScript!"; </script>`
Secured src	`<script src="java.js"> </script>`

```
&lt;p id=&quot;demo&quot;&gt;&lt;/p&gt;  &lt;script&gt;                                        Secured
document.getElementById(&quot;demo&quot;).innerHTML = &quot;Hello JavaScript!&quot;;;
&lt;/script&gt;
&lt;script src=&quot;java.js&quot;&gt; &lt;/script&gt;                                          Secured src
```

Fig. 4 Example of the escaping method

Even though, this is the most secured method out of the all named methods, it is still not completely secured. It is possible to input malicious strings into some places. One example of this is when URL is provided by the user input, like in this example:

    document.querySelector('a').href = userInput

While *href* gets the assigned value, it is automatically encoded so it becomes nothing more than the value of the attribute. This does not prevent the hacker from entering a URL that begins with *"javascript:"*. By clicking on the link any JavaScript code embedded in the URL will be executed. Encoding is also not an appropriate solution if the final goal is to prevent the user to set the part of the code to the web page. An example that illustrates this is a user's profile page where a user can define a custom HTML. If this custom HTML is encoded, the profile page may consist only of the plain text. In such situations, encoding must be completed with a validation.

3 Program Code

index.php:
```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width">
    <meta name="description" content="Cross Site Scripting">
    <meta name="keywords" content="Cross Site Scripting, Milan Tuba, XSS">
    <meta name="author" content="Milan Tuba">
    <title>My Project | Welcome</title>
    <link rel="stylesheet" href="./css/style.css">
  </head>

  <body>
    <header>
      <div class="container">
        <div id="branding">
          <h1><span class="highlight">Cross Site Scripting</span> Project</h1>
        </div>

        <span class="menu-trigger">MENU</span>
            <div class="nav-menu">
                <ul class="clearfix">
```

```
          <li class="current"><a href="index.php">Home</a></li>
          <li><a href="xss_s.php">XSS Stored</a></li>
        </ul>
    </div>
      </header>
      <section id="showcase">
    <div class="container">
     <h1>Demonstration of XSS</h1>
      <p>Web site is made using HTML5, PHP, SQL and JavaScript. Project is made for demonstration of
Cross Site Scripting.</p>
    </div>
      </section>
  <section id="boxes">
    <div class="container">
     <div class="box">
      <img src="./img/logo_html.png">
      <h3>HTML5</h3>
      <p>HTML5 is a markup language used for structuring and presenting content on the World Wide Web.
It is the fifth and current version of the HTML standard.</p>
     </div>
     <div class="box">
      <img src="./img/logo_css.png">
      <h3>CSS3</h3>
      <p>CSS3 is the latest evolution of the Cascading Style Sheets language and aims at extending
CSS2.1. It brings a lot of long-awaited novelties, like rounded corners, shadows, gradients, ...</p>
     </div>
     <div class="box">
      <img src="./img/logo_php.png">
      <h3>PHP</h3>
      <p>PHP (recursive acronym for PHP: Hypertext Preprocessor) is a widely-used open source general-
purpose scripting language that is especially suited for web development and can be embedded into
HTML.</p>
     </div>
    </div>
  </section>
  </header>
  <footer>
   <p>Milan Tuba, University Singidunum, Faculty for Informatic and Computing, 2019</p>
  </footer>
 </body>
</html>
```

xss_s.php:

```
<!DOCTYPE html>
<html>
 <head>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width">
  <meta name="description" content="Cross Site Scripting">
  <meta name="keywords" content="Cross Site Scripting, Milan Tuba, XSS">
  <meta name="author" content="Milan Tuba">
  <title>My Project | XSS Stored</title>
  <link rel="stylesheet" href="./css/style.css">
 </head>
 <body>
  <header>
   <div class="container">
     <div id="branding">
      <h1><span class="highlight">Cross Site Scripting</span> Project</h1>
     </div>
     <span class="menu-trigger">MENU</span>
        <div class="nav-menu">
            <ul class="clearfix">
```

```html
    <li><a href="index.php">Home</a></li>
    <li class="current"><a href="xss_s.php">XSS Stored</a></li>
  </ul>
</div>
  </header>
<section id="main">
 <div class="container">
  <article id="main-col">
    <h1 class="page-title">Guestbook</h1>
            <form action="insert.php" method="post">
                <input type="radio" name="dif" value="low" checked> Low
                <input type="radio" name="dif" value="medium"> Medium
                <input type="radio" name="dif" value="hard"> Hard
                <input type="radio" name="dif" value="sec"> Secured <br>
                Name: <br><input type="text" name="fname" style="width: 200px;" /><br>
                Comment: <br><input type="text" name="comment" style="width: 500px; height:
100px;" /><br><br>
                <input type="submit" />
            </form>
            <br>
    <table>
     <tr><th class="num">Name</th><th>Comment</th></tr>
     <?php
      $conn = mysqli_connect("192.168.89.66","test2","","test2") or die("Error Connection!");
      $query = "SELECT * FROM guestbook";
      $result = mysqli_query($conn,$query);
      while ($row = mysqli_fetch_array($result)) {
       echo                  "<tr><td              style='text-align:              center;font-weight:
bold;'>".$row['name']."</td><td>".$row['comment']."</td></tr>";
      }
     ?>
    </table>
  </article>
  <aside id="sidebar">
   <div class="dark">
   <h3>About subject</h3>
   <p>
     Understanding  major  concepts  regarding  the  Cross Site Scripting,  such  as  the  theoretical
aspects  and
  practical
implementations of the preventing methods. Developing strong web applications using up-
to-date practices and tools.
   </p>
   </div>
   </aside>
  </div>
 </section>
 <footer>
  <p>Milan Tuba, University Singidunum, Faculty for Informatic and Computing, 2019</p>
 </footer>
</body>
</html>
```

insert.php:

```php
<?php
    $dbhost="192.168.89.66";
    $dbname="test2";
    $dbusername="test2";
    $dbpassword="";
    $message = trim( $_POST[ 'comment' ] );
    $name    = trim( $_POST[ 'fname' ] );
    $dif     = trim( $_POST[ 'dif' ] );
    if($dif=="low"){
```

```
        $conn = mysqli_connect("192.168.89.66","test2","","test2") or die("Error Connection!");
    $message = stripslashes( $message );
    $message = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $message ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $name = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $name ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $query = "INSERT INTO guestbook ( comment, name ) VALUES ( '$message', '$name' );";
    $result = mysqli_query($conn,  $query ) or die( '<pre>' . ((is_object($conn)) ? mysqli_error($conn) :
(($___mysqli_res = mysqli_connect_error()) ? $___mysqli_res : false)) . '</pre>' );
    header("Location:http://localhost/xss_s.php");
    exit;
    }
    else {if($dif=="medium"){
    $conn = mysqli_connect("192.168.89.66","test2","","test2") or die("Error Connection!");
    $message = str_replace( '<script>','', $message);
    $message = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $message ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $name = str_replace( '<script>','', $name );
    $name = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $name ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $query = "INSERT INTO guestbook ( comment, name ) VALUES ( '$message', '$name' );";
    $result = mysqli_query($conn,  $query ) or die( '<pre>' . ((is_object($conn)) ? mysqli_error($conn) :
(($___mysqli_res = mysqli_connect_error()) ? $___mysqli_res : false)) . '</pre>' );
    header("Location:http://localhost/xss_s.php");
    exit;
    }
    else {if($dif=="hard"){
    $conn = mysqli_connect("192.168.89.66","test2","","test2") or die("Error Connection!");
    $message = strip_tags(addslashes($message));
    $message = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $message ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $message = htmlspecialchars( $message );
    $name = strip_tags(addslashes($name));
    $name = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $name ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $name = htmlspecialchars( $name );
    $query = "INSERT INTO guestbook ( comment, name ) VALUES ( '$message', '$name' );";
    $result = mysqli_query($conn,  $query ) or die( '<pre>' . ((is_object($conn)) ? mysqli_error($conn) :
(($___mysqli_res = mysqli_connect_error()) ? $___mysqli_res : false)) . '</pre>' );
    header("Location:http://localhost/xss_s.php");
    exit;
    }
    else {if($dif=="sec"){
    $con = new PDO("mysql:host=$dbhost;dbname=$dbname", $dbusername, $dbpassword);
    $conn = mysqli_connect("192.168.89.66","test2","","test2") or die("Error Connection!");
    $message = stripslashes( $message );
    $message = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $message ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $message = htmlspecialchars( $message );
    $name = stripslashes( $name );
    $name = ((isset($conn) && is_object($conn)) ? mysqli_real_escape_string($conn,  $name ) :
((trigger_error("[MySQLConverterToo] Fix the mysql_escape_string() call! This code does not work.",
E_USER_ERROR)) ? "" : ""));
    $name = htmlspecialchars( $name );
    $data = $con->prepare( 'INSERT INTO guestbook ( comment, name ) VALUES ( :message, :name );' );
    $data->bindParam( ':message', $message, PDO::PARAM_STR );
```

```
    $data->bindParam( ':name', $name, PDO::PARAM_STR );
    $data->execute();
    header("Location:http://localhost/xss_s.php");
    exit;
    }}
    }
    }
?>
```

java.js:
```
document.write("Hello JavaScript! external source");
```

References

[1] V. Nithya, S. Lakshmana Pandian, and C. Malarvizhi. *A survey on detection and prevention of cross-site scripting attack.* International Journal of Security and Its Applications 9(3), 139-152, 2015.

[2] T. Hadpawat, and Dipesh Vaya. *Analysis of Prevention of XSS Attacks at Client Side.* International Journal of Computer Applications 173(10), 1-4, 2017.

[3] I. Hydara et al. *Current state of research on cross-site scripting (XSS)–A systematic literature review.* Information and Software Technology 58, 170-186, 2015.

[4] M. Johns. *Code-injection Vulnerabilities in Web Applications—Exemplified at Cross-site Scripting.* IT-Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik 53.5, 256-260, 2011.

[5] B. B. Gupta et al. *Cross-site scripting (XSS) abuse and defense: exploitation on several testing bed environments and its defense.* Journal of Information Privacy and Security 11.2, 118-136, 2015.

Milan III TUBA
Singidunum University
Faculty of Informatics and Computing
Danijelova 32, 11000 Belgrade
SERBIA
E-mail: milan.tuba.17@singimail.rs

Stefan ILIC
Singidunum University
Faculty of Informatics and Computing
Danijelova 32, 11000 Belgrade
SERBIA
E-mail: stefan.ilic.17@singimail.rs

Framework for digital image processing in frequency domain

Una Tuba, Ira Tuba

Abstract

Digital images are widely used in numerous areas such as medicine, agriculture, astronomy, etc. The need for various digital image processing algorithms is high and many of them were proposed in the last decades. In general, these algorithms can be divided into two major group, digital image processing in spatial and frequency domain. Image processing methods in spatial domain deals with pixel values while in frequency domain, image is analysed by looking the rate at which the pixel values are changed in spatial domain. Numerous image characteristics can be better detected in frequency domain than in spatial. Since this represents an important problem, we proposed a framework for digital image processing in frequency domain where several standard algorithms were implemented.

1 Introduction

Digital images have an important role in both, every day life and science. Nowadays, numerous sciences use applications that process digital images with aim to detect various details and information which was either impossible or very hard before. One of the examples where usage of digital images can look unexpected or unusual is agriculture. Some of the state-of-the-art methods for plant disease detection are based on analyzing digital images obtain by strategically placed surveillance cameras, cameras attached to the unmanned aerial vehicles, i.e. drones, satellites or similar. Other way for discovering plant disease is manual observation by experts which is both, expensive and rather slow. In addition, well designed application can detect diseases that can be invisible to the human eye since computer processing of digital images can detect rather small anomalies and changes since it is not relying on imperfect organ such as human eye [4, 7]. Digital image processing methods are also used for soil examination [5]. Another examples where digital images introduced rather revolutionary changes are medicine [1], astronomy [2], biology and ecology [6], etc.

Even though the purposes of use of digital images can be very different, processing tasks are rather common. In general, digital image processing techniques can be divided into three main categories: low, middle and high level processing methods. Low level algorithms manipulate with the image without the knowledge about what the image is and what objects. Low level image processing methods are usually used in pre-processing step for contrast or brightness adjustment, binarization, etc. Middle level methods are used for finding different features in image such as edges, contours, textures, etc. The goal of middle level image processing algorithms is to detect different objects of interest. High level image processing methods usually include artificial intelligence where the objects detected by middle level methods are recognized. These algorithms usually works on object level and includes machine learning or other artificial intelligence, i.e. extracted features are processed in order to give them a meaning. For example, middle level algorithms can detect faces by searching for circular regions with skin colored pixels. High level methods would connect faces with their names, i.e. recognize detected persons.

Digital image processing can be done in spatial or frequency domain. In spatial domain, processing is done by manipulation with pixel values while in frequency domain, frequency coefficients were changed

and analyzed. Frequency domain is rather important for signal processing since numerous features can be easier detected. Also, most of the image compression algorithms are done in frequency domain by using the fact that human eye is less sensitive to high frequencies, i.e. edges and noise [8].

In this paper we present a software framework for digital image processing in frequency domain. We implemented several standard filters along with one of the most used compression method, JPEG algorithm. Presented framework enables user to set parameters of filters and see the resulting image compared to the original.

The rest of the paper is organized as follows. Two-dimensional discrete transformation used for transforming an image to frequency domain is presented in Section 2. Implemented algorithms for digital image processing in frequency domain are defined and explained in Section 3. In Section 4 we presented our proposed software framework along with some simulation examples. At the end, conclusion and future work are given in Section 5.

2 Frequency domain

Digital image transformation to the frequency domain is done by two-dimensional discrete cosine transform (2D-DCT). Frequency coefficients are defined by the following equation:

$$D(u,v) = \sqrt{\frac{2}{N}}\sqrt{\frac{2}{M}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} C(i)C(j)I(i,j)$$
$$\cos[\frac{\pi * u}{2 * N} * (2i+1)] \cos[\frac{\pi * v}{2 * M} * (2i+1)], \tag{1}$$

where 2D-DCT is applied to the block of size $N \times M$ and $D(x,y)$ is DCT coefficient, $I(i,j)$ is intensity level of the pixel (i,j) while $C(i)$ is defined as:

$$C(i) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } i = 0 \\ 1 & otherwise \end{cases} \tag{2}$$

After applying 2D-DCT image is transform and instead of pixel values, frequency coefficients are obtained. The very first coefficient is called DC component and it represents the average intensity. The rest of coefficients are AC coefficients. Low frequencies represent slow changes between neighbor pixels which means uniform parts of the image such as grass, sky, water surfaces, etc. These coefficients are placed around DC component. Coefficient that are closer to the opposite corner of DC component, closer to the right lower corner, represent high frequencies. Large difference or fast changes in intensities between neighbor pixels are represented by high frequency coefficients. By analyzing high frequency coefficients, information about edges and noise can be obtained. Since both of these tasks represents rather important part of numerous image processing applications, various algorithms for edge detection and noise removal in frequency domain have been proposed in the past. Lower frequency analysis is used for creating efficient compression algorithms and also for detecting different features in an image.

Image in spatial domain can be obtained by performing inverse 2D-DCT. Since digital image is represented by integer numbers and by applying 2D-DCT real numbers will be obtained. By using inverse 2D-DCT, some rounding error will appear. Inverse two dimensional discrete cosine transform is defined by the following equation:

$$I(x,y) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} C(i)C(j)D(i,j)$$
$$\cos[\frac{\pi * i}{2 * N} * (2x+1)] \cos[\frac{\pi * y}{2 * M} * (2y+1)], \tag{3}$$

where the same notation as before was used.

3 Digital image processing in frequency domain

Image processing algorithms in frequency domain modify DCT coefficients calculated by 2D-DCT transform and than the image is obtained by applying inverse transform. Image processing in frequency domain steps are presented in Algorithm (1).
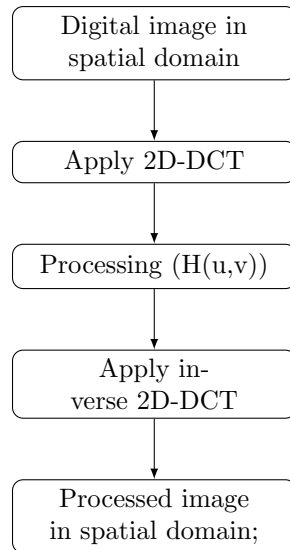
Algorithm 1: Image processing in frequency domain

Mathematical definition of filters in frequency domain is:

$$g(x, y) = IDCT[H(u, v)DCT(x, y)], \tag{4}$$

where IDCT is inverse 2D-DCT, $DCT(x, y)$ is two dimensional discrete cosine transform of the input image, $H(u, v)$ is a filter function and $g(x, y)$ is the filtered, i.e. output image.

One of the simplest filters in frequency domain is the one where the DC component is rejected, i.e. set to 0. Since DC component keeps the information about the average intensity, this filter will set average intensity to zero. The output image will be much darker compared to the input image. Having an average intensity equal to zero will result with some negative intensity values. Since digital image can not be represented by negative intensities, after applying inverse 2D-DCT, they are usually all set to zero.

Other well-known and widely used filters are lowpass and highpass filters. As it was explained before, low frequencies that are placed around DC component represent slow vary in intensity levels which occurs in uniform parts such as walls, cloudless sky and similar surfaces in image. Lowpass filter keeps low frequencies, i.e. lets them pass, while cutting high frequencies. Mathematically, lowpass filter keeps coefficients above some of the diagonals. If k is the number of *diagonal rows* that will be kept, them coefficients with indexes that satisfy the following condition are kept: $i + j < k$ where $i = 0, 1, 2, \ldots, N$ and $j = 0, 1, 2, \ldots, M$. The lowpass filter produced smoothed image since high frequencies, i.e. edges, are cut. Besides smoothing the edges, some types of noise will be removed.

The opposite filter to lowpass is highpass filter that keeps high frequencies and low frequency coefficients set to zero. This filter will keep sharp details but it will also reduce the contrast in the image. Highpass filter keeps coefficients at the positions (i, j) where $i + j > k$ and k is the number of *diagonal rows* that are kept.By applying appropriate highpass filter, edges can be detected.

Bandpass filter keeps the coefficients between two diagonals, i.e. frequency coefficients that are with indexes (i, j) and $k < i + j < l$. In this way $l - k$ rows starting from k^{th} diagonal will be kept. Since lowpass filter blur the image but remove the noise while highpass filter is reducing blurring but accentuates noise, the compromise is to use bandpass filter. By cutting the highest and the lowest frequencies the noise can be removed and the blur effect can be reduced comparing to usage of lowpass filter.

Unsharp mask and highboost filter represent processes that used for sharpening images. The process consists of several steps. The first step is to smooth the image and than to obtain edges by subtracting smoothed image from the original one. Obtained edges are than added to the original image. This is unsharp mask while highboost filter is obtained when the edges are multiplied by some constant before adding them to the original image.

In frequency domain, these two sharpening filters are defined in the following way. Mask, i.e. edges, are obtained as [3]:

$$g_{mask}(x,y) = f(x,y) - f_{LP}(x,y), \tag{5}$$

where

$$f_{LP}(x,y) = IDCT[H_{LP}(u,v)DCT(x,y)], \tag{6}$$

where $H_{LP}(u,v)$ is lowpass filter applied to image defined by $f(x,y)$. Image defined by $f_{LP}(x,y)$ represents smoothed image. Unsharp mask and highboost filter are defined as:

$$g(x,y) = f(x,y) + k * g_{mask}(x,y), \tag{7}$$

where for $k = 1$ unsharp mask is defined while for $k > 1$ is for highboost filter.

Besides these basic filters, frequency domain is used for image compression. One of the most used lossless compression algorithm (compression with losing some data irreversibly) is JPEG algorithm. The main compression in JPEG algorithm is obtained by reducing or completely removing some of the frequency coefficients. It was concluded that human eye is less sensitive to high frequencies so compression can be obtain by removing some of them or at least by reducing the size needed for their recording in memory. This step of modifying frequency coefficients in JPEG algorithm is called quantization step where matrix of coefficient is divided, element by element, by quantization matrix. By choosing appropriate elements of a quantization matrix good ratio quality to compression level can be obtained. We included in our framework JPEG algorithm where the recommended quantization matrices Q_i were used (i=1,2,...,100). The number in the quantization matrix name determine the ratio of compression level and the quality of the resulting image. Standard quantization matrix is Q_{50} and all other matrices are obtained based on it. The higher numbers represent higher quality but lower compression level while the lower number stand for higher compression and lower quality.
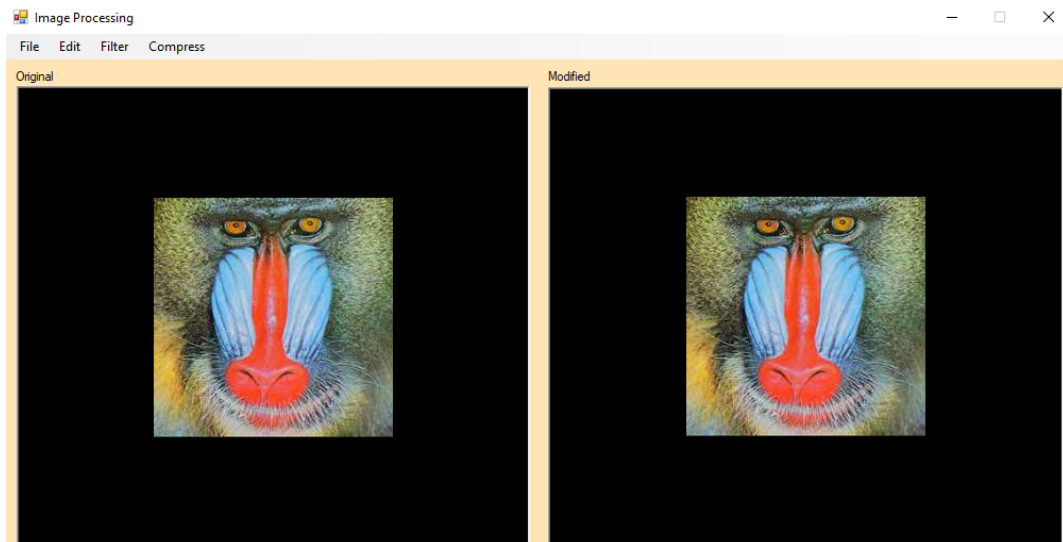

Figure 1: The proposed framework

4   Our proposed framework

In this paper we proposed a software framework for image processing in frequency domain. Filters described in the previous sections were implemented and all mentioned parameters are set by a user.

The proposed interface was implemented in Microsoft Visual C# 2017, NET Framework 4.7.1. The first thing that have to be done is to choose the image that should be processed. After opening an image, it will be shown on the left side of the window while the resulting image will be shown on the right side (Fig. 1).

For highpass and lowpass filter parameter k, i.e. the number of diagonal rows that will be kept should be defined. We set the default value to be 1. In Fig. 2 are shown results of applying lowpass and highpass filter, respectively.
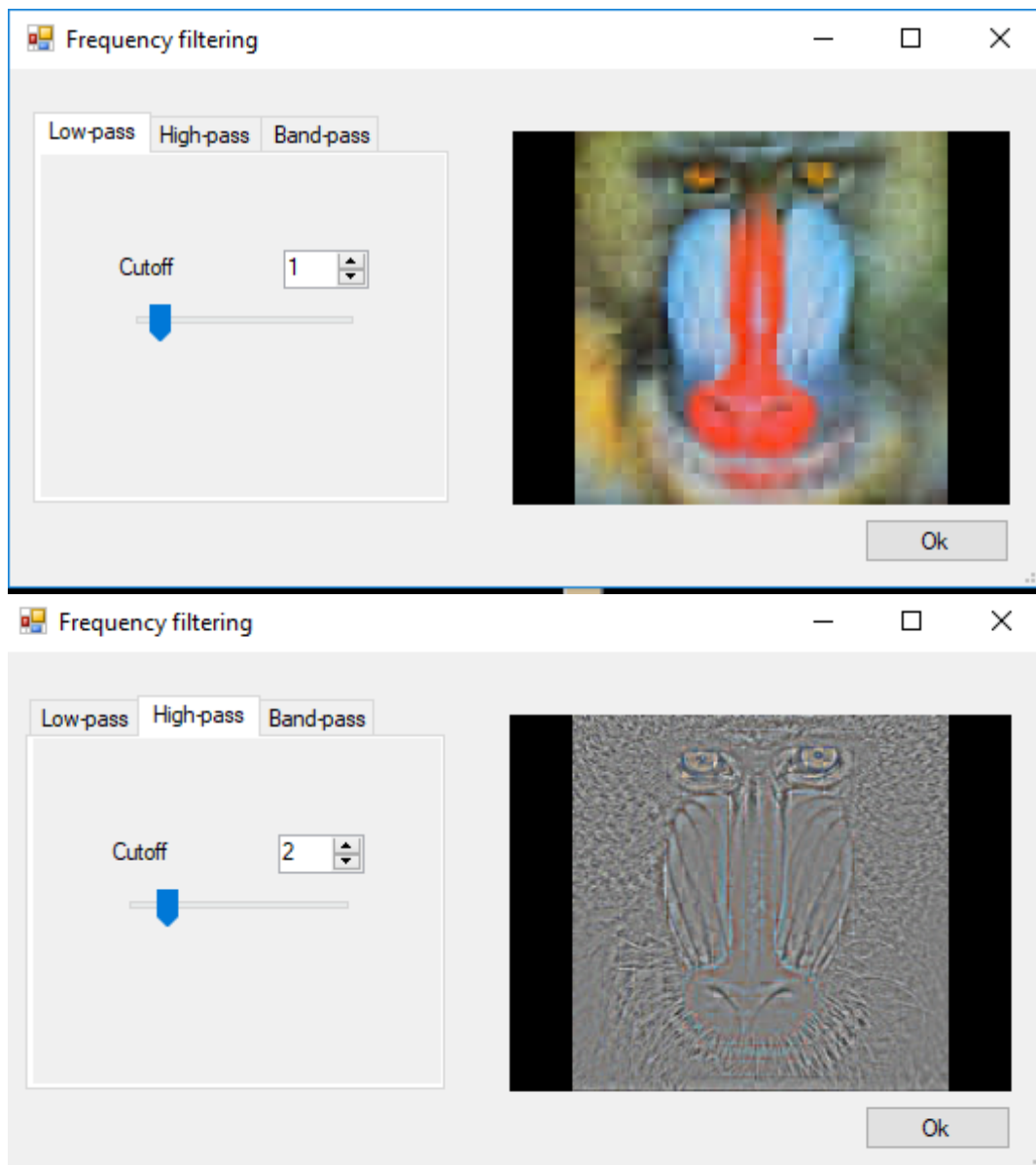

Figure 2: The results of applying lowpass (up) and highpass (down) filter

Besides lowpass and highpass filter, we implemented bandpass filter where the middle diagonals of

frequencies were saved. For bandpass filter, it is necessary to choose how many diagonals above and below the main diagonal will be saved. Bandpass filter is rather efficient for removing certain types of noise while reducing the blurring effect that is produced when low pass filter is used. The look of the proposed framework along with the resulting image is presented in Fig. 3.
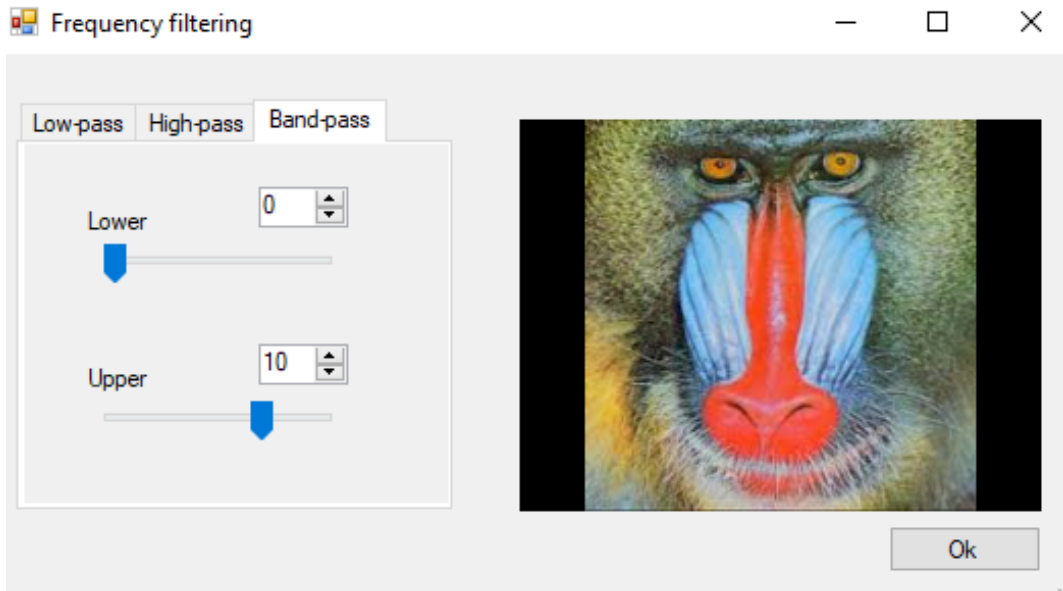

Figure 3: The results of applying bandpass filter

In our proposed framework for digital image processing in frequency domain, we included one of the simplest edge detection method described in the Section 2 where the high-boost filter was used. For edge detection by the presented algorithm, it is necessary to choose high-boost filter coefficient. Th resulting look of the proposed framework after applying edge detection method is presented in Fig. 4.
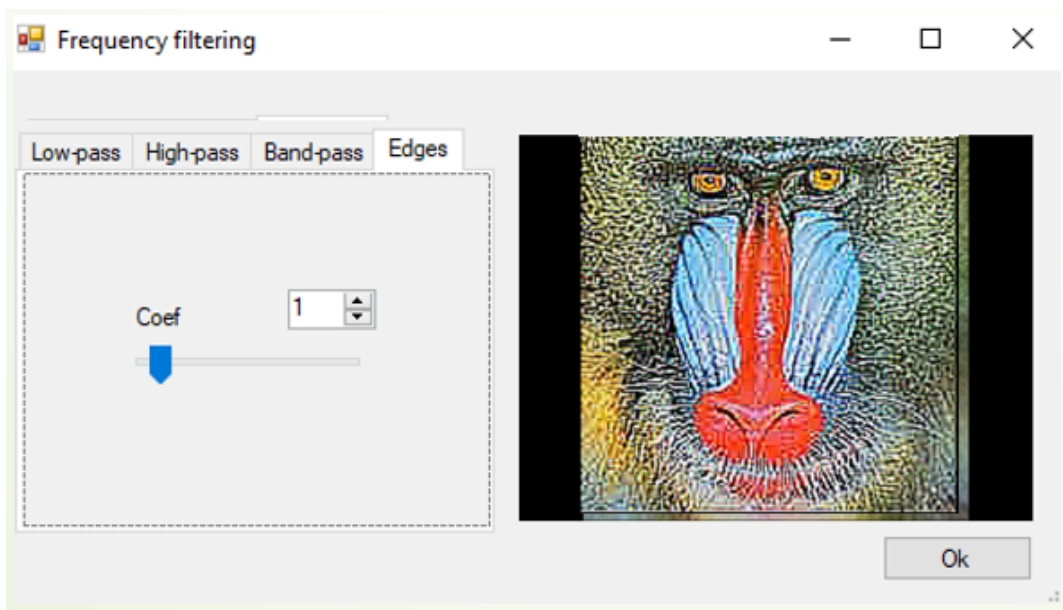

Figure 4: The results of applying edge detection algorithm

The last processing method included in this initial version of the framework is digital image compression. We implemented JPEG compression algorithm with an option to choose the quality/compression ratio. The default quantization matrix that is used is $Q_5 0$ matrix but the user can decide to use any other Q_i matrix. One of the plans for future work is to add an option for a user to choose each element of the quantization matrix and to allow him/her to customize compression result. The proposed framework for digital image compression in frequency domain by JPEG algorithm is presented in Fig. 5.
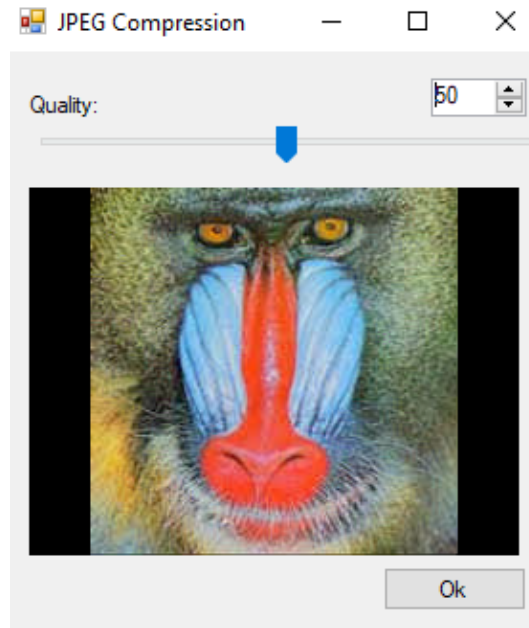

Figure 5: The results of applying JPEG compression algorithm with Q_{50} quantization matrix

5   Conclusion

Digital images have been widely used in numerous areas and digital image processing became one of the most important scientific fields. Numerous hidden features in a digital image can be captured in the frequency domain thus digital image processing in frequency domain is very important. In this paper we proposed a simple framework for manipulating with digital images in frequency domain. We implemented some of the basic filters as well as some advanced methods of processing digital images in frequency domain. In future work it is planned to add more filters and to make the framework more flexible in a sense that user can choose almost all parameters of the each filter that is implemented.

References

[1] S. Aja-Fernandez, G. Vegas-Sanchez-Ferrero, and M. A. M. Fernández, "Soft thresholding for medical image segmentation," in *IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*.   IEEE, 2010, pp. 4752–4755.

[2] K. A. Collins, J. F. Kielkopf, K. G. Stassun, and F. V. Hessman, "Astroimagej: image processing and photometric extraction for ultra-precise astronomical light curves," *The Astronomical Journal*, vol. 153, no. 2, p. 77, 2017.

[3] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[4] S. D. Khirade and A. Patil, "Plant disease detection using image processing," in *International conference on computing communication control and automation*. IEEE, 2015, pp. 768–771.

[5] G. P. Petropoulos, G. Ireland, and B. Barrett, "Surface soil moisture retrievals from remote sensing: Current status, products & future trends," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 83, pp. 36–56, 2015.

[6] M. Shortis and E. H. D. Abdo, "A review of underwater stereo-image measurement for marine biology and ecology applications," in *Oceanography and marine biology*. CRC Press, 2016, pp. 269–304.

[7] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017.

[8] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*, 4th ed. United Kingdom: Cengage Learning, 2014.

UNA TUBA
Singidunum University 1
Faculty of Informatics and Computing
Danijelova 32, 11000 Belgrade
SERBIA
E-mail: *tuba.una@gmail.com*

IRA TUBA
Singidunum University
Faculty of Informatics and Computing
Danijelova 32, 11000 Belgrade
SERBIA
E-mail: *ira.tuba@gmail.com*

Online Platform for Agricultural Industry

Ana Vesic, Sava Lakicevic, Vuk Ignjatovic, Bojan Gutic, Dusan Dotlic, Andrej Micovic

Abstract

This project has the goal of filling a massive gap in the market of agriculture in Serbia, which is a large branch of the country‟s economy. The low state of popularity and the problems often associated with the industry is what this application intends to resolve. The project will help people put themselves on the map and place their products on the market, while creating a healthy environment for business development, event promotion and cooperation. The application follows current trends and needs for organic food and sustainable production, while steering away from animal abuse and corporations taking advantage of farmers. Furthermore, this project intends to reward people doing honest work and adapt to new demands from consumers, as well as facilitate interconnectedness of farmers, associations, companies and the government. This document contains the idea, functionalities, user interface, architecture, the group‟s motivation for this project, and the current progress.

1   Introduction

When given an opportunity to develop a project with the help and guidance of a mentor, our newly formed group decided to make something meaningful, something useful, something significant. Since agriculture is a large branch of Serbia‟s economy and government, and has an underdeveloped e-governance, in addition to other problems, we have taken it upon ourselves to do research in this area. We studied what steps were taken in this department by the Serbian government, as well as what other countries have done with an introduction of e-governance. Consequently, we decided to construct a project that would contain a uniform set of solutions, which are in accordance with the technological development plan in Serbia. These solutions are implemented within a single easy-to-use web platform, complemented by an elegant and clean user interface. This platform is for all members of the agricultural industry; manufacturers, big or small, farmers or other large businesses, distributors, people providing transportation services, buyers, and the government.

There will be two sides to this project. The first one is focused around the activities of our users: farmers, transporters, resellers, associations, buyers, etc. The application through its functionalities will make communication and trade between affiliates easier and connect members of the application with each other. The platform will also serve as a facilitator for event organization and promotion, such as fairs and festivals, which will in turn help in the preservation of local traditions, while at the same time enabling self-promotion and supporting local associations. A calendar on the page will serve as a reminder for any kind of activities on the website, while also having the option to serve as a personal reminder for anything else. The second part of the project is focused around e-

governance and education. When it comes to education, we will be providing links to studies and academic research in the field, certain tips, a detailed weather forecast, recommended fertilizers and promoting competitions, fairs and seminars. In addition, we will provide information on how to gain certain certificates and licenses, how to register a new business and new mechanization, or other types of related documentation. In turn, we are going to use the provided information by the users to update and create new databases of ownership, mechanization, amount of production, used mechanization and owned licenses and certificates, while creating a detailed map of Serbian agricultural associations and farmland.

2   Background

Taking into account the world‟s trends and stage of technological development in government‟s systems, and the largest branch of Serbia‟s economy, we‟ve decided to contribute to, and improve its agriculture. Furthermore, we examined where digitalization is most necessary, and compared its current state with EU‟s requirements for the ongoing process of integration. While constructing the solution, we were researching what has been done so far in the neighboring countries, the world, and especially, in the EU. Previous to that, we carefully analyzed Serbia‟s concrete needs, considering all of the stakeholders, and took into consideration all of the political trends and plans, steps made, and the future plans. Throughout this research process, parts of our solutions that we had already constructed proved to be good, other parts needed improvement, and we discovered we need to add new parts that were not taken into consideration before. Acknowledging the importance of this question, we decided to include it in our final solution.

There are many governmental and non-governmental organizations, international and domestic institutions active on the topic of ICT in agriculture all over the world. Whether it is promoting ideas and development, giving support in the media, providing funds, organizing events, or providing the concrete solutions and the required infrastructure for their implementation.

Placing the manufacturers and their products on the market, promoting and enabling easier trade, better price regulation and regional cooperation in the field of agriculture is something that the international community, including the UN and the EU, is strongly encouraging. Accordingly, Food and Agriculture Organization of the United Nations has developed an online forum, e-Agriculture, where the good practices and information are exchanged from all over the world. Besides online, they organize regional forums and conduct research of comparative practices. They define E-agriculture as: designing, developing and applying innovative ways to use information and communication technologies (ICTs) in the rural domain, with a primary focus on agriculture and food [1] which is something that we aim to realize. They establish standards and norms to be achieved, and try to make and contribute to the national strategies in agriculture while including ICT in them. Furthermore, they promote fair and equal treatment and space for development to all of the stakeholders in the industry. Common Agricultural Policy of the EU is a big influencer, policy and standard setter in Europe, but it also strongly affects the wider international scene. CAP was founded in 1962 and some of its goals today are: „“…, enhancing the viability and competitiveness of all types of agriculture, and promoting innovative farm technologies... animal welfare and risk management in agriculture... promoting social inclusion, poverty reduction and economic development in rural areas.” [2]. This project is one of the largest in the EU and it receives very high amount of funds. It has a great impact on shaping policies, not only in the countries of the European Union, but also in those that are in the process of integration, and even wider. These two earlier mentioned organizations cooperate very tightly, making this development possible. Membership, integrations and plans for integrations to the EU are crucial to the development of ICT in agriculture. Not only do they fund and give support for the improvement, but they set standards which are ought to be followed in order to be competitive on the market. “This is explained by the legal obligation set by the EU to operate an integrated information system in relation with the Common Agricultural Policy.” The conclusions and instructions that the Food and

Agricultural Organization has set are: "The farmer should be in the center of the strategy, which should not be „forgotten" during the process driven by the governmental authorities and other stakeholders isolated from the „end users"… the strategy should be planned and implemented to ensure a win-win for farmer and government (and other stakeholders)…The pace of high-tech innovation is increasing in agriculture… for this reason we will need a functional e-strategy on national and international level that may be the pillar of the future farming regulations and sustainable smart agriculture"[3]. Among these, there is another important goal to mention: "Stimulate collaboration and knowledge sharing via online communities of practice, including existing regional networks…". Therefore, in order for Member States of the EU to obtain the funds and subsidies within the CAP, it is required to conduct an Integrated Administration and Control System (IACS), as well as to fulfill certain conditions such as having computerized and interconnected databases which are used to receive and process aid applications and respective data. Thus it provides for: a unique identification system for farmers, an identification system covering all agricultural areas called land parcel identification system, an identification system for payment entitlements, a system for identification and registration of animals [3].

One thing that we noticed that is of great importance, but given little attention coupled by the present lack of knowledge, is business and risk management in agriculture. It is only recently that in Serbia, agriculture is considered as a part of entrepreneurship, but the very business management is not something that is given much attention to. The world trend of providing management for businesses, including risk calculations, trends, predictions, statistics, etc., in order to improve and make decision-making easier should be incorporated in provided aid and education to the Serbian manufacturers, alongside with IoT technologies with the goal to improve the quality and efficiency of food production. However, it is important to note that the IoT technologies are something that is evolving in Serbia, with startups that have had decent success. Besides this one, we noted more of the concrete solutions from around the world. In India, Slovenia, Moldova and other countries governments" web portals are implemented with different aspects of agricultural information that can be needed and required to access them easily. From concrete products, to existing risks in quality of soil, weather conditions in different parts of the country, to various statistics, land registration, etc., all of the information are provided to the government from manufacturers, experts and other stakeholders in order to be published online later. Moldova has implemented a project aiming to provide greater exposure to the manufacturers, by placing the products on the online market for different traders. It has been incorporated into the larger national extension service website, and it provides a portfolio of important information to traders and producers. Besides that, there is also a web portal for domestic and international trade. In Hungary similar project was done, only it provides relevant information to easily find producers and healthy agricultural in the neighborhood. Furthermore, there are numerous projects considering advisory services – e.g. weather predictions and adjusted tips, among other educational, and advice-giving features. In Slovakia the service provides information for professionals, the public and farmers on livestock, milk analysis including breeding information system. Projects considering e-government are various as well. Agricultural Information System in Turkey records all areas of Turkish agriculture aiming to produce agricultural statistics in all of the fields, as well as planning the production and usage of subsidies in order for them to be made effective. Georgia has developed the agro-portal, where the farmers can see and get all services from the same place as the Ministry and its member agencies. The portal for obtaining various datasets for viewing and downloading, with the purpose to use relevant data for applying for the funds is developed in Slovenia. In countries like Kyrgyzstan the registration of parcels and updating relevant databases is needed, so the portal for real estate registration was brought to the farmers. There are many other projects considering risk management, providing various information for markets, innovation systems etc.

Since 2005 in Serbia a lot of effort has been put into digitalization, especially in areas of e-governance. However, very little has been done so far in the field of agriculture, except for the

projects providing information about available subsidies or funds, and practical tips for the farmers. Other than that, all that was realized was done on a local scale, with no big-scale impact. Agricultural industry makes 10% of value added in percent of the GDP, one quarter of Serbia‟s population is a part of it, and 57% of land area is dedicated to agricultural purposes. It is the biggest, most important and most perspective branch of Serbia‟s economy and export. However, the problems are present in the matter of export, market and price stability, lack of machinery and proper education (especially considering new and young farmers, coming from cities with the great support of states‟ policies, and the education considering business management), hard or impossible access to needed information and statistics, outdated databases, underdevelopment of rural areas and local communities, decrease in popularity and interest in the agricultural industry. It is very important to emphasize the importance of communication between all of the stakeholders in agriculture, including the government. By enhancing the quality of exchanging important information, providing fast and easy document obtaining, business related registration and application for relevant subsidies and programs, not only time will be saved, but it would be much more efficient, economically sustainable, and the development in this industry would be much easier to achieve. By enabling subjects for easy business and obligations transactions the motivation rises and improvement is a necessary consequence. One of Serbia‟s recent policies was the revival of farmers‟ associations. Therefore, developing rural areas in social and infrastructural aspects by connecting people and providing the content and conditions for growth on the local level, is very necessary. Taking into consideration all of the said above, and the fact that Serbia is a member of United Nations, has participated in FOA‟s forums, and is in the process of EU integrations, we developed further functionalities, providing a uniformed solution.

3   Functionalities and Implementation

The web platform for this project was developed as a 3-tier client-server application, which is a common choice in application development today, and is consisting of presentation tier (frontend), business logic tier (backend), and data tier (database). The frontend part of the functionalities developed for this project and described in the document is nearly completely ready and were developed using Bootstrap, and later adding the JavaScript code. The backend, however, is currently a work in progress and is being developed with the help and guidance of our mentor. The backend is developed as a RESTful web service in the Java Spring framework, and is completed for user management and posting. This way of developing the backend was chosen, because it is easy to maintain and is highly scalable. We used Spring Boot used for backend development as a good provider of support for enterprise web application development. The API was implemented in a traditional 3-layers way: REST controller, Service and Data Access (DAO). REST controllers are providing responses in JSON, as well as the business logic. REST Endpoints were defined using the @RequestMapping annotation, and the basic CRUD operations are implemented: GET (return user by ID or all users), POST (create new user), PUT (update user); other controllers are similarly implemented. In the figure below is the data model for this web platform.
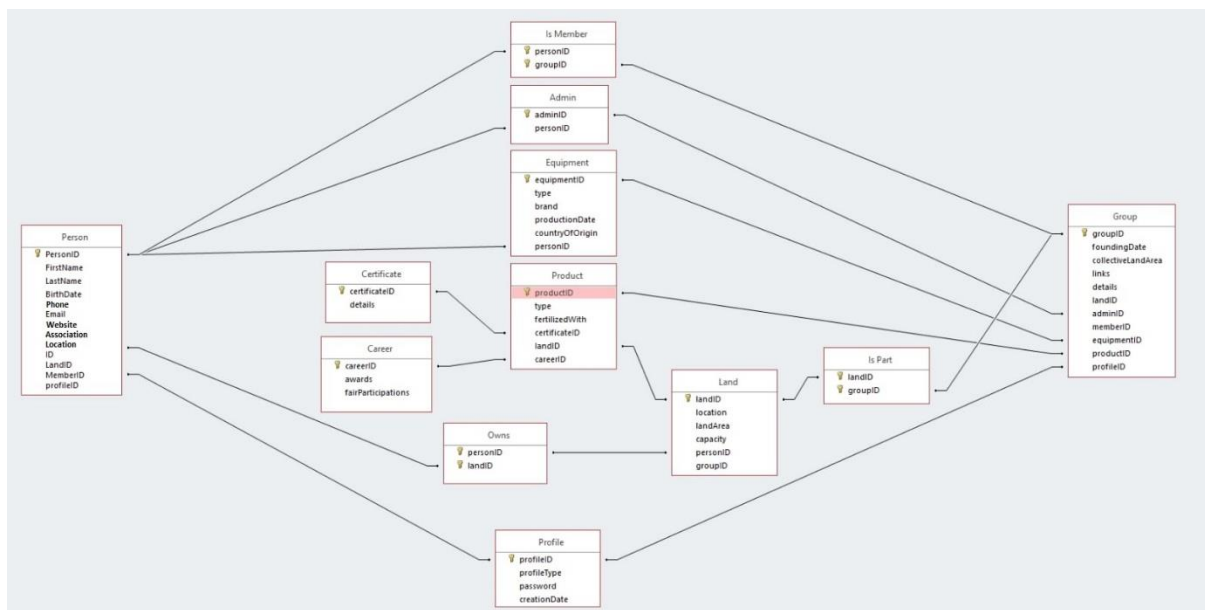
Fig. 1. Database Model

3.1 Registration and User Management

A person needs to have an account on the platform in order to access all of the information and functionalities on the platform. When registering on the platform the user will need to enter some personal and business information. First of all, the user selects a type of profile, an individual, an association, a company, or another type of a group. The personal information will include name, surname, place and date of birth and an e-mail. The business information, on the other hand, will include location, production, which association he is a member of, as well as owned certificates and licenses. The information that the user provides will be visible on his profile. However, some of the information will need to be reviewed and verified by an admin or a government body.

Fig. 2. Login Page

3.2 Visit profile, connect and join

First of all, when viewing a profile, its information will be displayed, as well as whether the information is verified or not; this will serve as a prevention mechanism for potential scams and misinformation. The posts and recent activities of other members on the platform that a user is

connected to or of associates he is a member of will be displayed on the home page. Notifications about the posts from associates and groups will be automatically on, while the option for notifications of other members is a choice. These and other features can be adjusted at any time based on the user"s preferences. If a user has selected to connect to another profile, or a is member of a group, the latest activities and new updates of these will show up on the user"s home page. When requesting to join a group, the admin will receive a notification with the option of accepting or rejecting the application. This process can be adjusted to made automatic, without the need to wait for the approval of an admin. When trying to contact another user, a request will be sent to the other person, who can in turn accept or decline the message. In addition, people will have the ability to block other users, or report them for harassment, spam, deceitful business or bots and hacked accounts.
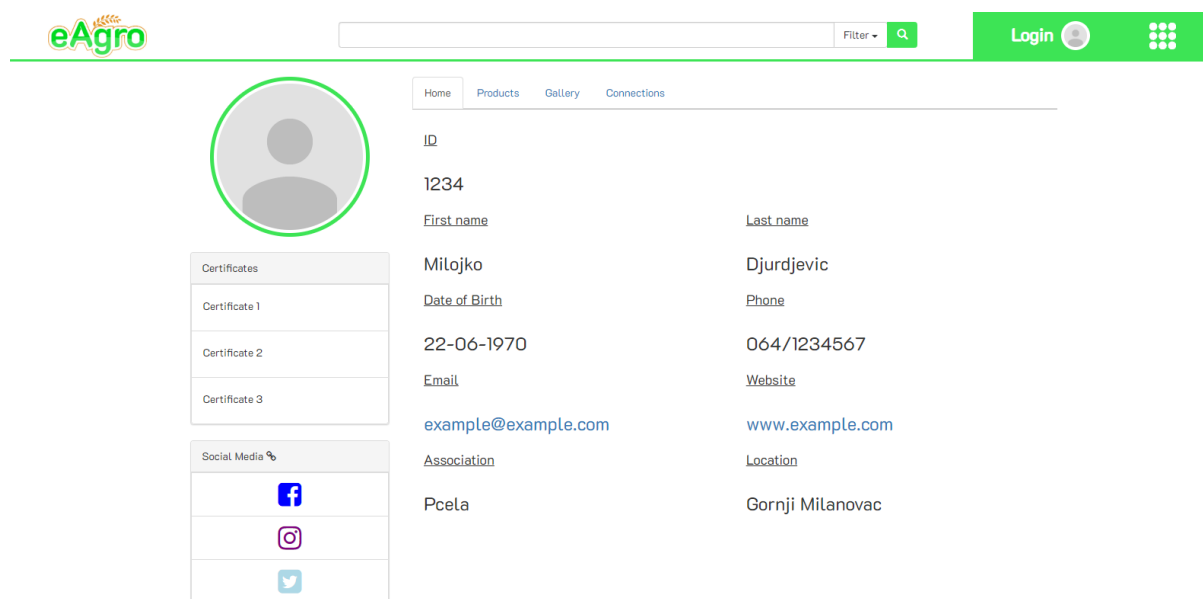
Fig. 3. Example of View Profile Page

3.3 Post

When posting on the platform, users will have an opportunity to use many different provided options in the application. Organizing and promoting events such as fairs, festivals, drink and food tastings, will be made easier on the platform. In addition to promoting events, the user will have an option to promote his own activities, services, as well as buy, sell or rent offers; these can be adapted to be for products, transportation services, equipment, fertilizers, fruit arching, and farmland or parcels. The posts can also be used for recruitment of employees, which is particularly useful for seasonal workers, like harvesting. People can also collect feedback and opinions from other members through polls; for example, bouncing ideas off of others, which product is better, optimal types of fertilizers, success of an event, etc.
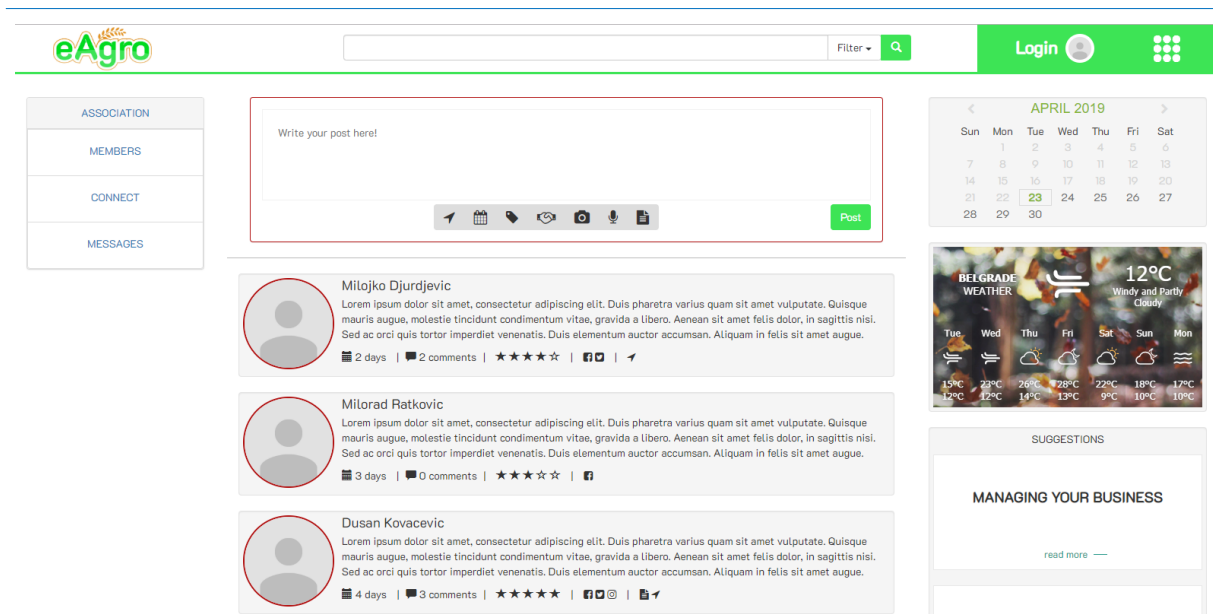
Fig. 4. Home Page

3.4   Home Page and Search Bar

The home page will contain recent updates, new activities and posts from other accounts that the user is connected to, or groups and associations he is a member of. In addition to scrolling through posts, the user will have a calendar and a weather forecast on his home page. The calendar will include activities on the platform, which will serve as a reminder for any kind of organized events that he had marked as interesting, or that he intends to attend. However, the calendar can also be used for any kind of personal activities off the platform. The weather forecast, next to the calendar, is an important feature for any manufacturer, which will be provided in a collaboration with certain meteorologists. On top of the home page, a search bar is located, with which anything that is on the platform can be found. When conducting a search, the user will have the option of filtering his search results, for example, by posts, people, groups, location, events, certificates, etc.
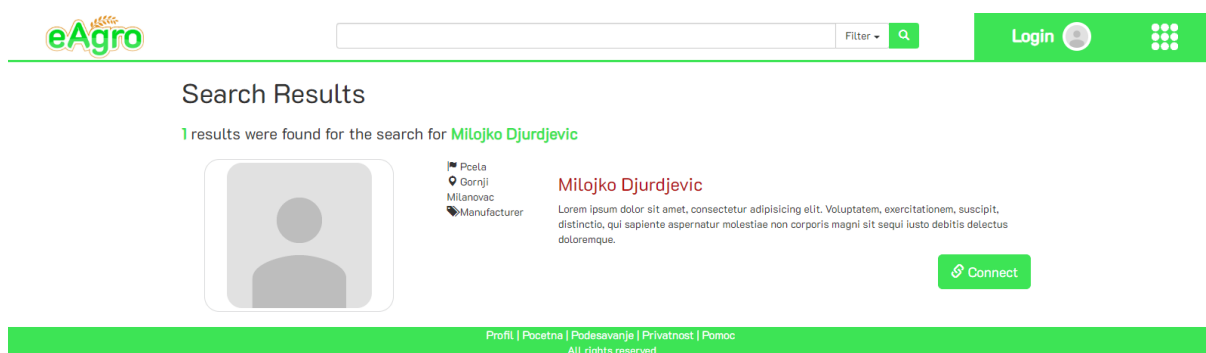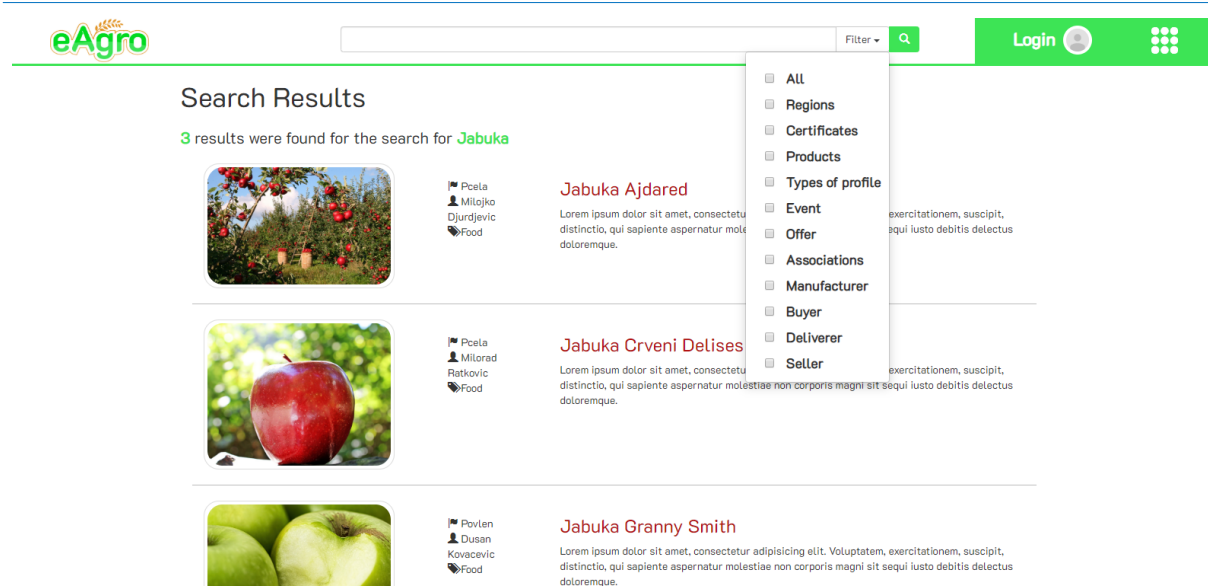

Fig. 5. Search Results Page, Example 1

Fig. 6. Search Results Page, Example 2

3.5   Education

On the home page, on the right side there will be links and suggestions provided for upcoming or trending educational materials. The functionality can be found in the main menu as well. It will lead the user to a new page where the materials will be divided into sections according to the topic, and if the course or lecture is provided online or live. Tips and advice will have their own section on the page as well. Educational materials will vary from tips how to behave in a certain weather situation, what is the best seed to plant and when to plant it, to business and risk management, etc. At the bottom of the page a post option will be enabled, in order for people to easily organize and promote educational events in their local communities. This feature will be available only to certain users (associations, local authorities). They can post an event, or make an open call for the lecture.
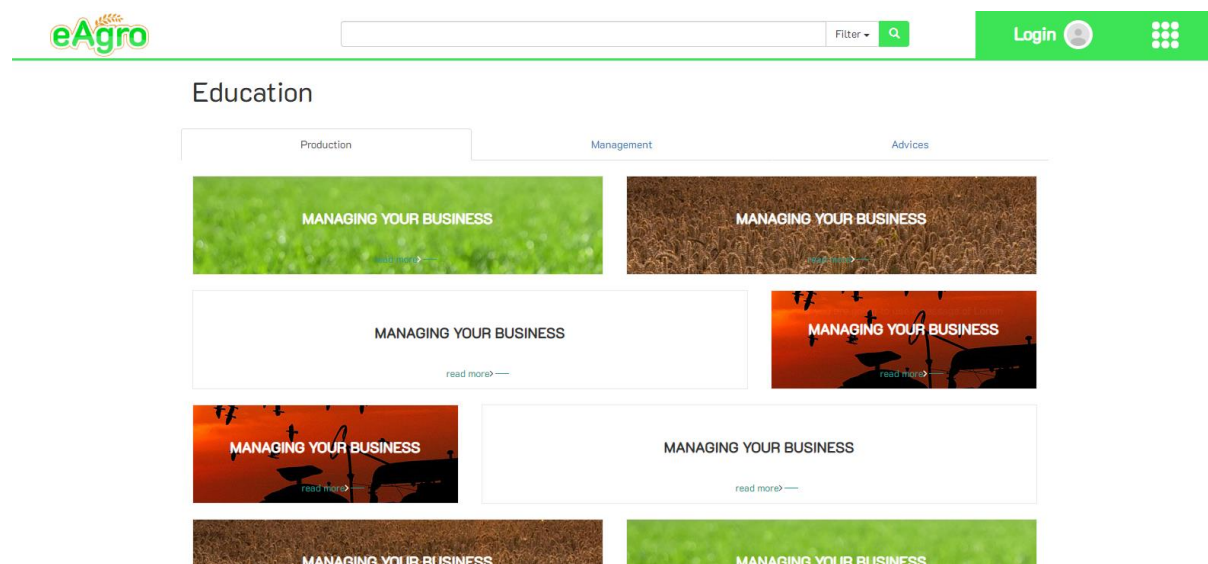

Fig. 7. Education Page with Suggested Courses and Tips

3.6   E-governance

E-governance is a set of features, as follows: registration of machinery, products, business, associations, obtaining documents such as certificates, various types of verifications, other legal and

non-legal documents, those containing information of their business etc., applying to funds, subsidies and other programs, submitting various applications, reporting problems (epidemic, problems in a local community etc.) When registering, by filling the form, the governments databases will be automatically updated. An important feature is locating themselves on the map.

Fig. 8. eGorernance Page

3.7   Settings

The page for adjusting settings will contain standard categories for the user to optimize his experience on the platform. The settings will contain priority of connections and groups, home page, privacy, notifications, login and profile information, and blocking. With these the user will be able to adjust the way he receives notifications, how many and from whom, as well as what information appears on his home page. The user will have control on how much of his information is displayed when somebody visits his profile, who can view his activities, who can contact him. The user will be able to add and edit his profile information, for example, to add new earned certificates, or licenses, or change any data. Certain changes to the profile will require further verification, just like when registering on the platform. The user will have the option of removing or adding new accounts to his block list, and mute or prevent notifications or requests to message from others.

Fig. 9. Create and Edit Profile Page

227

4   Conclusion

The rest of the functionalities are going to be fully implemented in the future. In addition, the platform has a great potential have other ideas and services added to it, which is a great advantage. Another advantage is achieving accessibility to solutions to current problems in Serbia, while creating a great environment for business, formation of new associations, and the popularization and further growth of agriculture. Current trends for organic require information about where and how products are made. The end product of all this is also satisfying new consumer demands, while benefiting all members of the industry. However, our application highly relies on a thriving community, as well as people coming on board and embracing this new idea. Convincing some older generations to use the platform may be problematic, even though we intend to provide tutorials on how to fully utilize all that the application has to offer. Furthermore, as first-year students, we lack experience and knowledge when dealing with already existing and future challenges, some of which we probably cannot even predict or expect.

References

[1] http://www.fao.org
[2] ec.europa.eu/agriculture/rural-development-2014-2020_en, on 23.3.2019.
[3] FAO. *Status of Implementation of e-Agriculture in Central and Eastern Europe and Central Asia - Insights from selected countries in Europe and Central Asia ,* Budapest, 52 pp, 2018.
[4] Isabella Rodriguez y Baena, Clare Pedrick, Sophie Treinen, Charmaine Wilkerson, *Stat SMS Gateway: Improving animal health through Information and Communication Technologies,* FAO, 2017

Ana VESIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: ana.vesic.18@singimail.rs

Sava LAKIĆEVIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: sava.lakicevic.18@singimail.rs

Vuk IGNJATOVIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: vuk.ignjatovic.18@singimail.rs

Bojan GUTIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: bojan.gutic.18@singimail.rs

Dusan DOTLIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: dusan.dotlic.18@singimail.rs

Andrej MIĆOVIĆ
Faculty of Technical Sciences, Singidunum University
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: andrej.micovic.18@singimail.rs

Jewelry design studio information system

Yuliya Vlasova

Abstract

This article discusses some aspects of the jewelry design studio information system. The article relevance is due to the apparent contradiction between the perspectivity of such a direction as jewelry design for individual orders and the shortage of software for automating the design studios activities, taking into account the peculiarities of jewelry design creative process. According to the results of the review of ready-made software solutions, it was concluded that it was necessary to develop an original information system. The author investigated the specificity of the subject area - jewelry design, analyzed the features of the jewelry designer's search activity during the work on an individual order. As a result, a data model was developed and a software application was created.

1 Introduction

Modern society places new demands on the market, the trend for the individualization of goods and services is growing. The consumer expects product originality, the opportunity to emphasize his style and individuality. As a rule, the value of the finished product in the jewelry design field is recognized by everyone as an intellectual activity result. At the same time, intermediate results - the results of a search and analytical activities, for example, are not given such importance as ready-made designer products. Nevertheless, the new product creation activity is preceded, in essence, by the research activity, because before creating something new, it is necessary to study everything that was created before in the future product concept framework.

At each stage there are intermediate results, so even the process of selecting and analyzing already existing analogs, creating collections of files, structured according to very different criteria, is already a valuable result. For the design studio knowledge to be correctly formed, stored, used, processed and replenished - an information system is needed, aimed primarily at working with information relating to the subject area and reflecting the features of creative activity.

Most jewelry design organizations use software products (for 3D modeling, graphics editors, etc.) while improving the 3D printing process broadens the scope for developing and manufacturing unique custom-made products. At the same time, the information processes themselves, which take place at the stage of product design development, as a rule, are not properly automated and are poorly reflected

in information systems. There is a wide choice of software to automate a variety of tasks; most of the business processes have already been automated. Nevertheless, there is a shortage of ready-made solutions in the market, focused primarily on the product design development process, which takes into account not only the order execution managing process but also the organization's knowledge management, informational support for search and analytical designer activities. Therefore, the development of specialized jewelry design information systems today is extremely important.

2 Jewelry design studio information system

2.1 Available jewelry design studio information systems overview

What does the ready-made software market for the automation of the jewelry design studio offer?
During the review of software (Table 1), which could be adapted to the tasks of jewelry design,
software products existing on the Russian market are divided into several areas:
- specialized (industry) solutions for the jewelry industry,
- software solutions for customer relationship management (CRM),
- software for project management,
- solutions for the automation of museum activities.

Title	Purpose of the software product and its description
Software solutions for the jewelry industry from "YuvelirSoft" [1] and "IT-Kostroma" [2]	Specialized software products (YuvelirSoft: Jewelry Production Management, YuvelirSoft: Jewelry Trading House, IT-K: Jewelry Company Management, IT-K: Jewelry Production, IT-K: Jewelry Trade Management, IT -K: Jeweler workshop and so on.) are developed taking into account jewelry production specifics, trade and special requirements for documentation and reporting. Contain jewelry classifiers, but they are not quite suitable for jewelry design studio information systems. The solutions are aimed at managing jewelry production and jewelry trade, keeping product records and raw materials, generating special reports, but do not provide information support for search and analytical activities.
1C: CRM PROF. Rev. 3.0 [3]	Support of business processes of the company per the CRM concept, including purchases, sales, marketing, service, etc. The CRM concept is suitable for working with a large customer number and is ineffective when working with customers in the segment of exclusive products "to order", does not allow for a comprehensive jewelry product description.
1C: PM Project Management PROF [3]	Automation of project management, project programs and project portfolios in enterprises and organizations. The program reflects the progress of work on the project, the allocation resources, and documentation storage, but does not allow for the product description and their search by various parameters.
1C: Museum [3]	Museum activities automation. Accounting and compilation of museum object scientific descriptions, registration of restoration works, search for objects, exhibition planning, excursion activities planning, reporting. It contains a "museum catalog" classifier that describes jewelry as objects of decorative and applied arts but does not reflect the product design developing process.

Table 1: Jewelry design studio Russian software review

The studied software is mainly aimed at the accounting of products, orders or customers, as well as the production management or marketing activities management, so it only solves some of the tasks.
None of the software products the groups can provide a ready-made solution for informatization of specific processes of a jewelry design studio that performs individual orders.
They do not fully solve such urgent problems as:
- collection and analysis of requirements;
- information support in the process of direct client communication and future jewelry product details discussing;
- search for necessary information, it's storage and processing;
- development of the concept, discussions with the client of the jewelry product preliminary sketches;
- amendments and design of the final (approved by the client) jewelry product design.
In this situation, it is advisable to develop an individual software solution that would meet the demands for the creative processes automation and contain all the functionality necessary for the studio to work effectively in the jewelry design developing process.

2.2 Data modeling taking into account jewelry design search activity features

The process of information searching [4] often occurs simultaneously with information analysis. The selected material can be analyzed again, changed and supplemented, and there can be many such iterations during the work on an individual order.
Information search is most often carried out in two cases:
- search for information in the process of direct communication with the client;
- search for information already in the course of independent work on an individual jewelry product.
Turning to the designer for the unique product development, the client rarely has an exact idea of what exactly he wants. The designer needs to understand the tastes and requests of the client as quickly and accurately as possible to correctly determine the range of tasks and the direction of work. Often the client's communication with the designer begins with a request to view the results of the studio's work, i.e. need to demonstrate the design of their jewelry. The client wants to see the jewelry studio products relating to the direction of interest and meet certain requirements. Using the information system as a knowledge base can help the designer to demonstrate the diversity of jewelry, to draw the client's attention to some original solution. Each client needs an individual approach, so the implementation of search functions will help to work effectively with knowledge in the subject area and improve the interaction and understanding between the client and the designer.
After the acquaintance with the client has passed, the requirements have been clarified and the concept of the future jewelry product has been defined, the designer continues to analyze the information received. It should be immediately noted that the search for materials necessary for the design development always begins with the study of the design studio already created in this field, as well as with the study of information from jewelry reference books. After that, naturally, the search is conducted in other sources to replenish the knowledge base.
The developed information system should solve the main tasks in the work of the jewelry design studio [5]: to support the process of research and creativity, preservation, use, and accumulation of knowledge; the accounting of the results of intellectual activity (both intermediate and final), as well as the accounting of jewelry products (completed orders), should be carried out.
The specificity of the jewelry design domain dictates the need to include in the data model a variety of different criteria for jewelry describing, and each criterion, in turn, can be accompanied by reference information.

The presented data model fragment (Figure 1) demonstrates how information about jewelry is stored (using the description of jewelry inserts in an item as an example), information about files, collections, and collection sets. It also shows how the jewelry reference books are stored, containing not only the jewelry characteristics but also the corresponding collection sets, correlated with the values in the jewelry reference book.
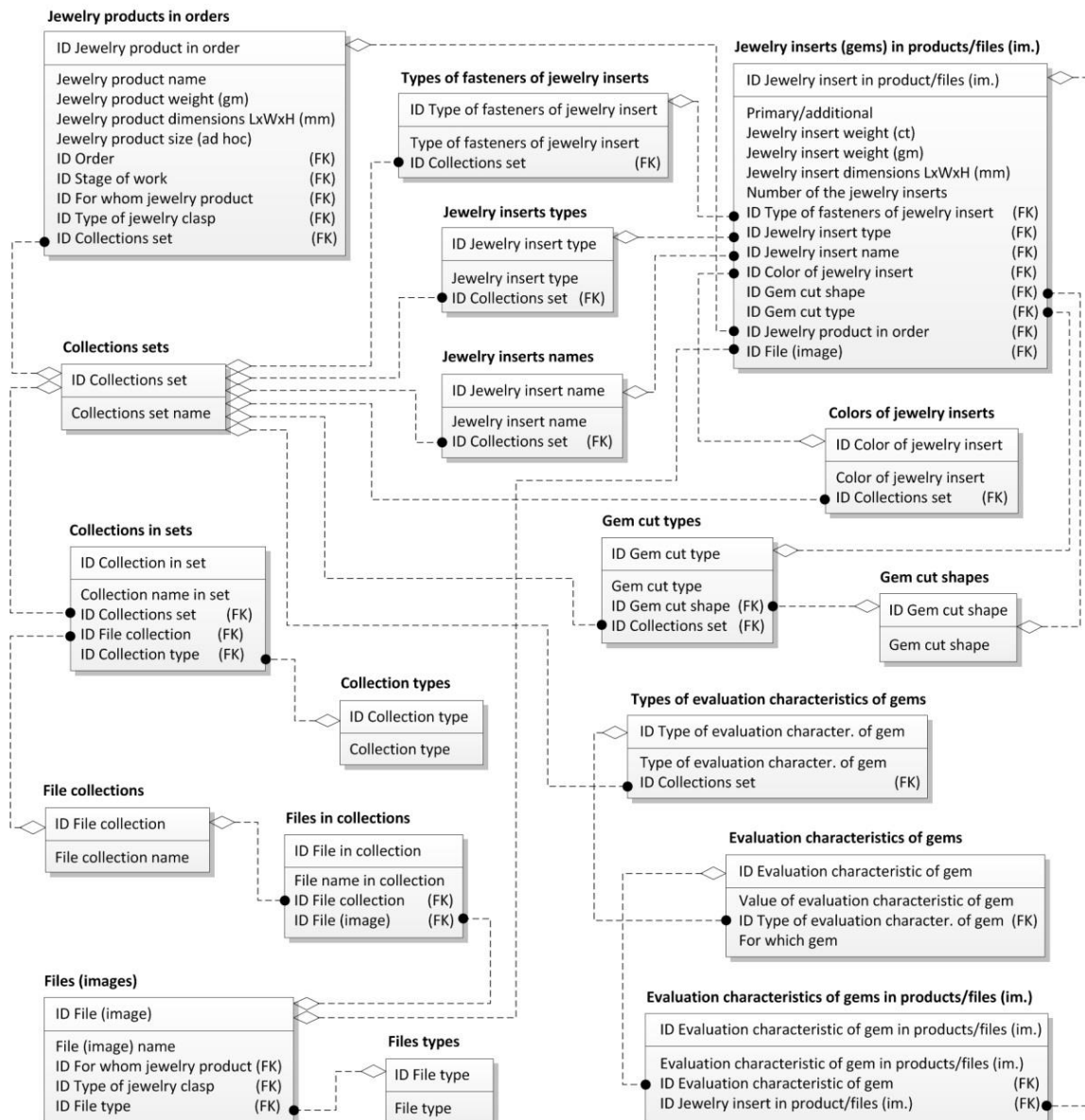
Fig. 1: Fragment of a logical data model

Description of jewelry products is the semantic center of the entire data model in the jewelry design subject area. This structure of the data model is due to several reasons: the need to fully describe the jewelry piece and the course of work on it to account for the jewelry product and the accuracy of its manufacture according to clearly defined parameters; the desire to effectively use the intellectual activity results, which were formed during the jewelry product development.

The difficulty in jewelry product description creating is that jewelry can be made from a combination of different metals, using several jewelry techniques, which can have several different jewelry inserts (primary or additional). A product can have a complex description of the type of product, several different themes and styles can be attached to the products, the jewelry product has several common descriptive characteristics, such as complexity of manufacture, shape, formal composition (for example, symmetrical or asymmetric), etc. And most importantly, for each jewelry product, you can attach a set of collections, in which files related to the developing process of jewelry design are stored in a structured form.

The main table in the presented fragment of the data model is "Jewelry products in orders". This table contains information about the product name, weight, dimensions, special size (for example, ring size), the order in which the jewelry product is located and the stage of work on this product. You can specify who the product is intended for (female, men's, etc.), the type of fastener and several other characteristics.

A jewelry product can have several inserts. It will be indicated that the insert is primary or additional. The weight of the jewelry insert, its dimensions, and several jewelry inserts will also be indicated.

 For the description of the insert, the corresponding values will be selected from the reference tables: "Types of fasteners of jewelry inserts", "Jewelry inserts types", "Jewelry inserts names", "Colors of jewelry inserts", "Gem cut shapes", "Gem cut types" (round brilliant)). The table "Evaluation characteristics of gems in products/files (images)" contains data of insertion characteristics (color grade, clarity grade, cut grade, etc.), these values correspond to the jewelry insert in the product.

For the description of the file loaded into the database (for example, images of any product), the "Jewelry inserts (gems) in products/files (images)" table can also be used, thus you can describe both real products and "virtual" ones (i.e. write the product image, stored in a file). The table "Files (images)" indicates the file name, file type, to whom the product is intended, the type of jewelry insert fasteners, etc. Files can be stored as separate objects with their descriptions, but most often they are part of collections. When uploading files to the system, it is possible to describe each file by a variety of parameters. It is also possible to create and store independent collections of files. Any collection can be designed and described as belonging to a specific set with an indication of the collection type. A set of collections can relate to the jewelry product, thereby expanding the description of this product and allowing you to save the results of the design development process. The "Files in Collections" table allows you to associate a specific file with a collection of files, and a collection of files using the "Collections in sets" table retrieves the collection type value and collections set. Sets from the "Collections sets" table can refer to the product description from the "Jewelry products in orders" table or from reference tables. Also, sets are used to store jewelry reference books information.

To effectively search, you need to describe in detail the files available in the information system, form collections of these files, and collect the collections into sets. The system provides a search for both in the sets for the jewelry products, the sets for the jewelry reference books, and for the individual files described, in this way you can select one, several or all sources for the search.

2.3 Information system implementation

To work on the information system project for a jewelry design studio, Microsoft.NET technology was chosen [6]. The technology includes the .NET Framework, a platform for developing and running applications. The Microsoft ADO.NET environment is part of the .NET Framework library that allows you to create applications for managing data in structured file or server repositories. To develop the database, it was decided to choose the MS SQL Server relational database management system [7, 8]. MS SQL Server provides for building client-server applications, accessing data from any application

developed using Microsoft .NET technology and Visual Studio. For convenient access and work with the database, a software application "Jewelry Design Studio" was developed using MS Visual Studio. Figure 2 shows the menu structure of the design bureau information system.

"Jewelry reference books" section contains the subsections: "Types of jewelry", "Jewelry materials", "Jewelry techniques", "Design features", "Jewelry product styles" and "Jewelry products themes". Jewelry reference books contain basic reference information on the subject area of jewelry design and are an important part of the developed information system. The section "Types of jewelry" includes the subsection "General information" and subsections with background information about the main types of jewelry, structured by purpose: "Personal jewelry", "Wardrobe items", "Writing accessories", "Watch accessories", "Souvenirs" and others. The section "Jewelry materials" includes subsections containing reference information about the main types of jewelry materials, divided into two main groups: "Metals" and "Jewelry inserts". Subsection "Metals" includes "Types of metals", "Metal samples", etc. The section "Jewelry techniques" includes information on various jewelry techniques and types of metal processing. The "Design features" section includes information on the types of fasteners (fasteners, locks) for various types of products, as well as on the types of fasteners of jewelry inserts.

In "Orders" section, there are subsections "Orders in progress", "Completed orders", "All orders", "Order search" and "Add order". In "Jewelry products" section, there are subsections "Jewelry products in progress", "Finished jewelry products", "All jewelry products", "Jewelry product search" and "Add jewelry product". In the menu of the "Clients" section, there are subsections "Active clients", "Inactive clients", "All clients", "Client search" and "Add client". "File Collections" section contains the subsections "Sets of collections", "File collections", "Files", "File search" and "Add file". "Reference books" section contains subsections "Types of services", "Services prices", "Prices on the jewelry materials", "Employees" and "Positions". In "Documents" section, there are subsections "Forms of documents", and "Reports". The subsection "Forms of documents" gives access to documents of various types and forms for creating new documents (technical assignment for 3D-modeling, work order for production, etc.). "System Information" section in the "User's manual", "System information", and "Developer information" subsections contain reference information about the system.
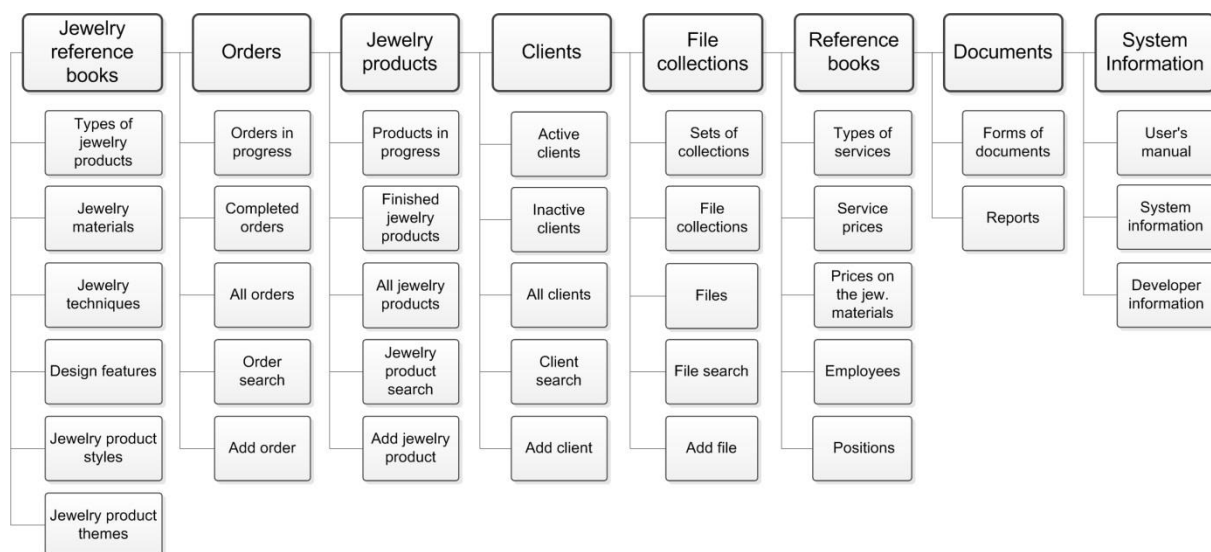

Fig. 2: Menu structure of the design studio information system.

An important part of the application is the "Jewelry product card", which allows you to conveniently organize effective work with the jewelry product, store all the necessary information about the product, display the results obtained in the design development process, and work with collections and documents. In the upper part of the window, there is general information about the product and buttons, with which you can go to the order, return to the jewelry product list, turn on card editing mode and save the changes. The "Jewelry product card" tab contains the tabs "General information about the jewelry product", "Metals and jewelry technology in the jewelry product", "Jewelry inserts in the product", "Jewelry product description", "Collections set for jewelry product" and "Appearance and documents for the jewelry product". In the main field of the tab "Collections set for jewelry product" (Figure 3), collections of files are displayed.
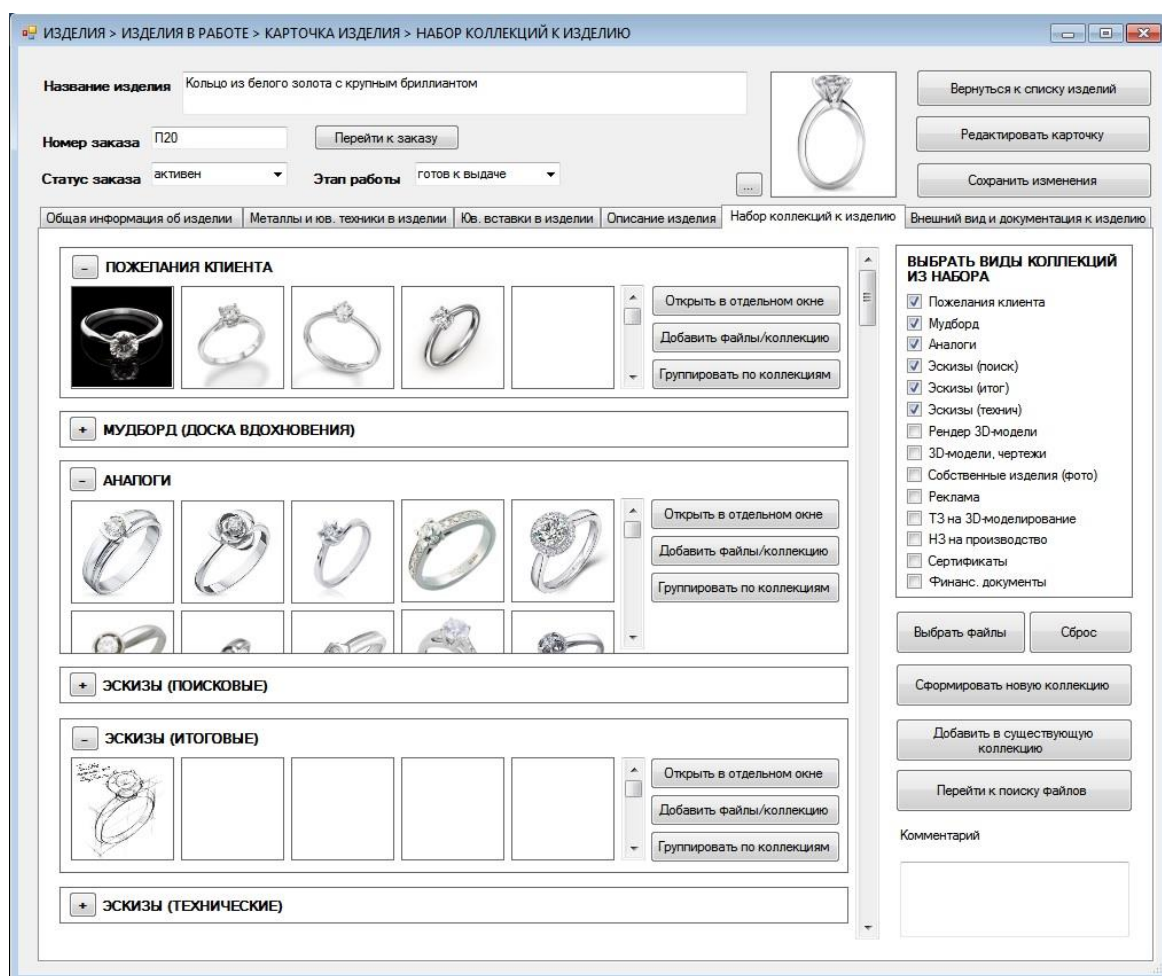
Fig. 3: Window of the tab "Collections set for jewelry product" ("Jewelry product card")

Which collections related to this jewelry product will be shown depends on the configuration of the list of types of collections (the necessary types of collections should be noted). Here it is conveniently organized to work with files and collections, it is possible to select files for generating new collections, add files, and also go to the file search window.

The window of one of the most popular types of gem cut – "Kr-57 (round brilliant)" from the reference book "Gem cut types" demonstrates the use of collections sets in organizing reference information in the system (Figure 4).
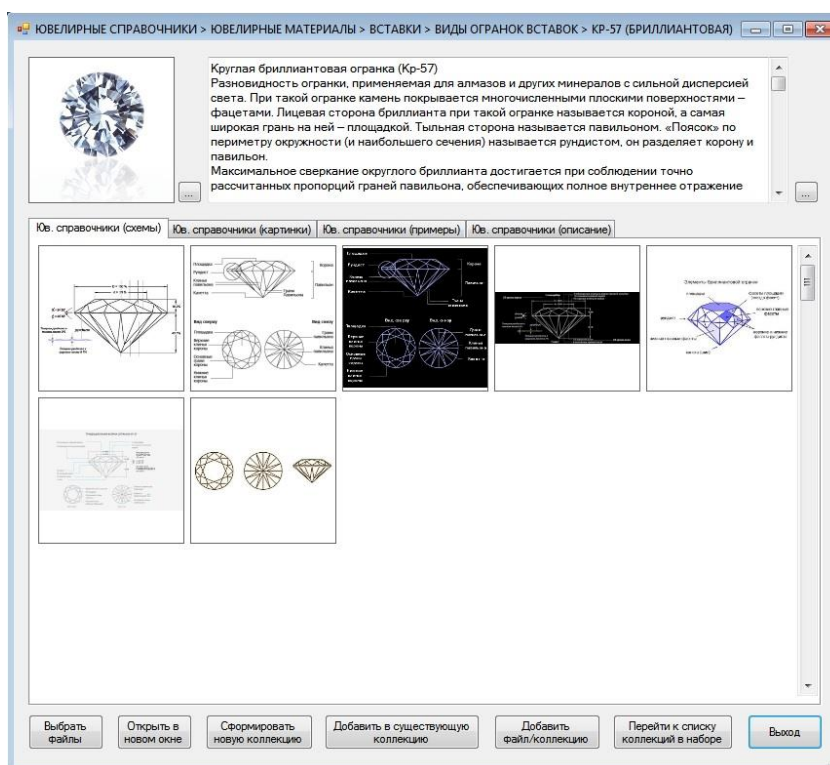
Fig. 4: Window "Kr-57 (round brilliant)" (Jewelry reference book "Gem cut types")

At the top of the window is general information on this type of gem cut. In the main field, you can select one of the four tabs to view files for a jewelry reference book, grouped by collections of a certain type, included in the set.

The description of the file added to the system is carried out in the "File description" (Figure 5).

Fig. 5: Window "File description"

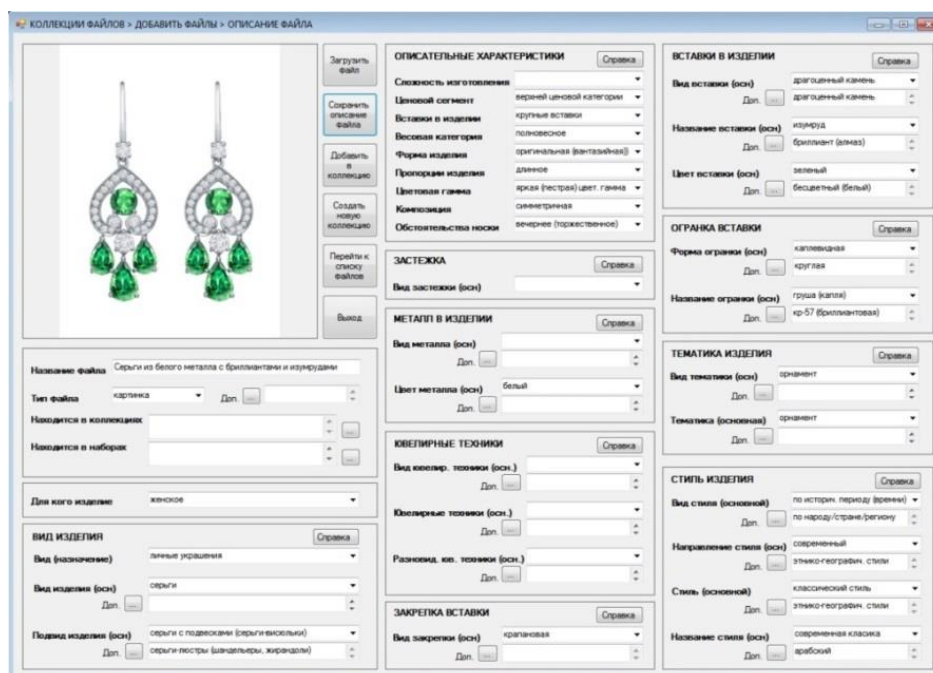The image of the jewelry product that we see when downloading the file can be described by choosing the values of the same characteristics as in the description of the jewelry product in the order.

We now turn to the direct search for files by the specified parameters. The window "File search" (Figure 6) of the software application contains the main descriptive characteristics of jewelry products, which have already been discussed above. In most cases, you can choose a basic value for each search parameter, as well as several additional ones. When setting up the search parameters, you can also select which types of collections will be searched. The search is organized in such a way as to give the most accurate result on request, but some subjectivity in the description of files and products is a fact to be considered, but it is precisely this individualization of the knowledge base that reflects the features of a rather abstract and weakly structured sphere of creativity.
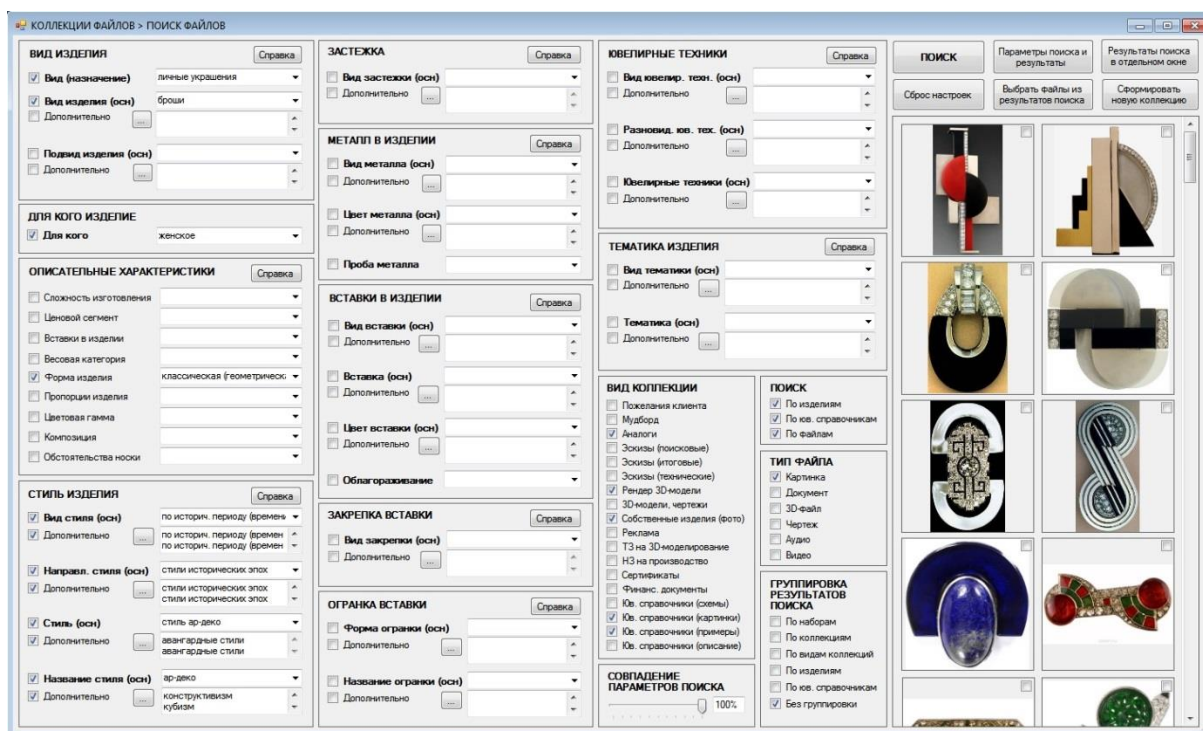
Fig. 6: Window "File search"

In this example, we have modeled a situation in which it is necessary to find images of an Art Deco brooch. In the search parameters, in this case, we note the type of jewelry product (by purpose) - personal jewelry, the type of jewelry product - brooches. We note that we need a women's jewelry product, and the form of the product is classical (geometric). We are looking for a product in the styles of historical epochs (a type of style - according to the historical period), Art Deco style, but additionally, we also choose avant-garde styles - constructivism and cubism. In the types of collections, we note, for example, analogs (that is, we are looking for alternatives in other collections that could have been compiled earlier when working with similar products), 3D model visualization files, photographs of finished jewelry products of the design studio, as well as jewelry reference books (pictures and examples). We select a search in sets of collections relating to jewelry products in orders and jewelry reference books, as well as a search for individual files (with a corresponding description

present). In the file types, we select - picture, do not use a grouping of search results. Click the "Search" button and see the results in a small area of the right side of the window "File search".

The preview is convenient because we can immediately see if something was found in the database and whether the result is acceptable to us. You can influence the search result by choosing the measure of the match of the selected search parameters, as well as by changing the search parameters and the selected values. For easy orientation in the search process, there are "Help" buttons on the window form (when you click the button, a window with brief reference information will appear on the value selected), and also when you select additional values, a window opens in which you can select values list and for everyone can see a little reference.

At this stage, you can already select the files of interest and create a new collection, but for convenient viewing, you can open the search results in a separate window (Figure 7).
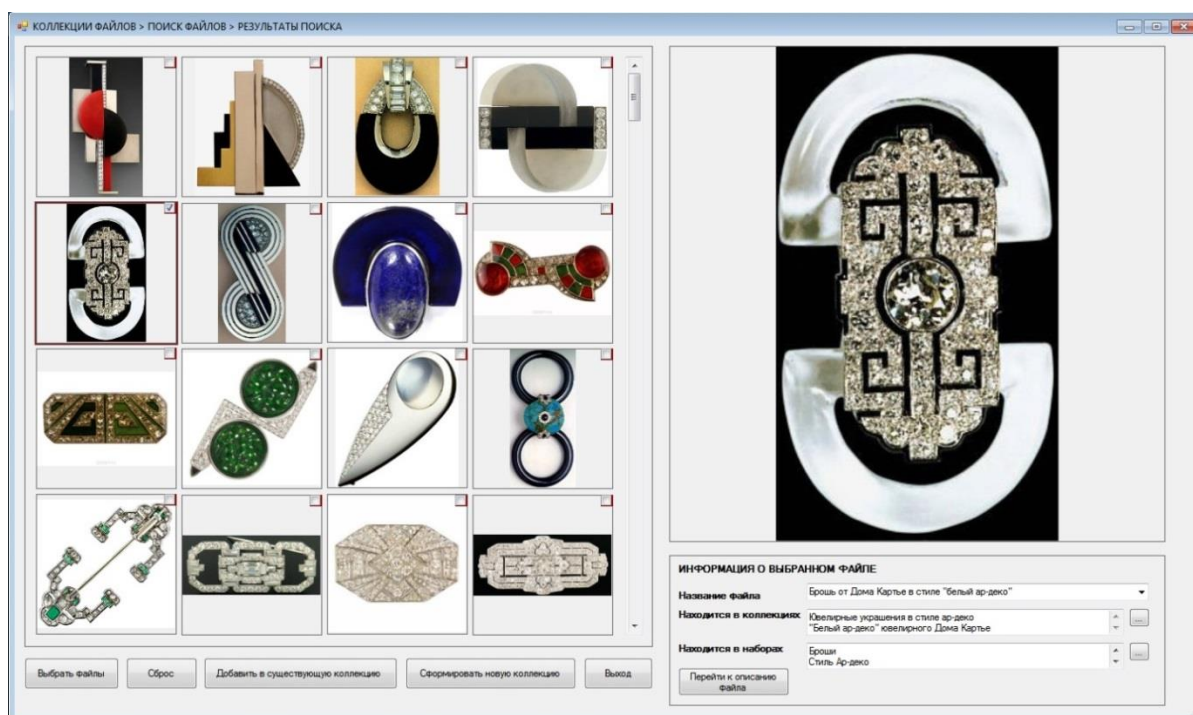
Fig. 7: Window "File Search Results"

This allows you to view all the files found on request, select the file of interest, see it in large format and get information about it (including in which collection and sets it is stored). You can go to the window form of the description of this file. And, of course, implemented the ability to select the desired files and create from them a new collection or add to an existing collection.

These examples show that the detail of the description of jewelry products and files (images), as well as the ability to search for a variety of different parameters, helps to obtain a more correct result.

The user interface of the application for the jewelry design studio has a convenient menu that reflects the peculiarities of work on the jewelry and contains the window forms most needed for the designer's work. Window forms are designed taking into account the specifics of the sphere of jewelry design and the features of information processes that accompany the development of jewelry product design. The application allows you to automate the search, analytical and creative processes that are an integral part of the design activity.

3   Conclusion

In the field of jewelry design, we see a clear contradiction - with the demand for the direction of product design for individual orders and the need to automate the activities of design studios, the ready software market is characterized by a lack of information systems reflecting the features of creative activity.

The development of an information system for a jewelry design studio is an urgent task, but rather non-trivial. A study of the jewelry design domain was carried out, the features of the search activity were analyzed during the work on an individual order, and a method for implementing the search function was proposed, as well as a way to organize complex descriptions of jewelry. As a result, a logical model of a relational database was developed and a software application for a jewelry design studio was created. For this information system, the use of the principles of collections and sets is provided, which allows you to organize a convenient search and work with information, i.e. solve some of the specific jewelry design problems.

The development of a specialized information system for a jewelry design studio continues and, possibly, during the work on the project, the use of other types of DBMS will be considered and other ways of solving specific jewelry design problems will be found.

References

[1] *YuvelirSoft* URL: https://uvelirsoft.ru/programs/
[2] *IT-Kostroma* URL: http:// www.it-kostroma.com/soft/
[3] *Otraslevye i specializirovannye resheniya 1S:Predpriyatie [Industry and specialized solutions 1C:Enterprise]* URL: https://solutions.1c.ru/catalog/
[4] Kravchenko Ju.A. *Zadachi semanticheskogo poiska, klassifikacii, strukturizacii i integracii informacii v kontekste problem upravlenija znanijami [The tasks of semantic search, classification, structuring and integration of information in the context of knowledge management problems]* // Izvestija JuFU. Tehnicheskie nauki. Razdel I. Analiz dannyh i upravlenie znanijami. 2016. № 7 (180) S. 5-18. (in Russian).
[5] Samarin Yu.N., Karpochev M.N. *Osnovnye zadachi avtomatizacii processov upravleniya informacionnymi potokami v studii dizajna [The main tasks of automating the management of information flows in a design studio]* //Vestnik Moskovskogo gosudarstvennogo universiteta pechati. 2016. №2. S.34-36. (in Russian).
[6] *NET Framework Guide* URL: https://docs.microsoft.com/en-us/dotnet/framework/
[7] Krigel A., Truhnov B. *SQL. Bibliya polzovatelya, 2-e-izdanie [SQL. User Bible, 2nd Edition]*. Per. s angl. M.: Vilyams, 2010. 752 s. (in Russian).
[8] *Rukovodstvo po SQL Server [SQL Server Guide]* URL: http://docs.microsoft.com/ru-ru/sql/sql-server/tutorials-for-sql-server-2016

Vlasova YULIYA
Volga Region State University of Service,
Department of Applied Informatics in Economics
4 Gagarin st., Togliatti, Samara region, 445677
RUSSIA
E-mail: jdvlasova@yandex.ru

LIST OF AUTHORS

Vasile ANGHELUȚĂ

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: angheluta.vasile@gmail.com

Andrei-Timote   ARDELEAN

Transilvania University of Brasov
Faculty of Mathematics and Informatics
Iuliu Maniu, no 50
ROMANIA
E-mail: timitoc@yahoo.com

Paula Ioana BARNA

Aurel Vlaicu University
Departament of Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: paula.barna@yahoo.com

Cooper BODARY

Cooper Bodary
College of Wooster
Computer Science Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: cbodary19@wooster.edu

Maria-Alexandra BOGDAN

Aurel Vlaicu University
Departament of Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail : bogdan.alexandra99@yahoo.com

Kristina BOZIC

Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
E-mail: kristina.bozic.16@singimail.rs

Tiberiu Cosmin BUJOR

Vasile Alecsandri University of Bacău
Faculty of Sciences
Bacău, Calea Mărășești 157
ROMANIA
E-mail: bujor.tiberiu@gmail.com

Dejan BULAJA

Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
E-mail: dejan.bulaja.16@singimail.rs

Iuliana-Maria BURUIANĂ

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: iuliana.buruiana@hotmail.com

Andreas COȘ

"Horea, Cloșca și Crișan" National College
B-dul 1 Decembrie 1918, No. 11, Alba Iulia
ROMÂNIA
E-mail: andreascos2002@gmail.com

Claudia CAUCA

Aurel Vlaicu University of Arad
Mathematics and Computer Science
Bd. Revoluției 77, 310295 Arad, Romania
ROMANIA
E-mail: cauca.claudia@yahoo.com

Alexandru CRĂCIUN

Iuliu Hațieganu University of Medicine and Pharmacy
Victor Babeș str. 8, Cluj-Napoca 400000
ROMÂNIA
E-mail: craciunalex98@yahoo.com

Valentin-Gabriel CRĂCIUN

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: gabriel.craciun@ulbsibiu.ro

Gheorghe-Cătălin CRIȘAN

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: crisan.gheorghecatalin@gmail.com

Aleksa CUK

Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
SERBIA
E-mail: aleksa.cuk.15@singimail.rs

Dušan DOTLIĆ

Singidunum University
Faculty of Technical Sciences,
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: dusan.dotlic.18@singimail.rs

Adoris Elian DORAN

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: elian.doran@ulbsibiu.ro

Răzvan Gheorghe FILEA   Samuel von Brukenthal National College
                             Sibiu, ROMANIA
                             E-mail: razvan.filea@gmail.com

Erika GOETZ   College of Wooster
                             Computer Sci. Dept.
                             308 E. University St.
                             Wooster, OH 44691
                             USA
                             E-mail: egoetz20@wooster.edu

Matei GRAURĂ   Lucian Blaga University of Sibiu
                             Faculty of Science - Informatics
                             Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
                             ROMANIA
                             E-mail: grauramatei@gmail.com

Pavlo GRYNKO   Simon Kuznets Kharkiv National University of Economics
                             Department of Higher Mathematics, Economic and
                             Mathematical Methods
                             Nauki avenue, 9-A, Kharkiv, 61166
                             UKRAINE
                             E-mail: malyarets@ukr.net

Bogdan-George GROS   Aurel Vlaicu University of Arad
                             Str. Elena Drăgoi, nr. 2,
                             310330, Arad
                             ROMANIA
                             E-mail: gros.bogdan@yahoo.com

Estera GROZA   Aurel Vlaicu University of Arad
                             Str. Elena Drăgoi, nr. 2, 310330, Arad
                             ROMANIA
                             E-mail: estygroza@yahoo.com

Bojan GUTIĆ   Faculty of Technical Sciences, Singidunum University
                             Software and Data Engineering
                             Danijelova 32
                             SERBIA
                             E-mail: bojan.gutic.18@singimail.rs

Svetoslav HADZHIIVANOV   University of Ruse
                             Department of Informatics and Information Technologies
                             8, Studentska str., Ruse 7017
                             BULGARIA
                             E-mail: warrolen@gmail.com

Felix HUSAC   Lucian Blaga University of Sibiu
                             Faculty of Science - Informatics
                             Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
                             ROMANIA
                             E-mail: husacfelix@gmail.com

Stefan ILIC   Singidunum University
                                   Faculty of Technical Sciences,
                                   Software and Data Engineering
                                   Danijelova 32
                                   SERBIA
                                   E-mail: stefan.ilic.17@singimail.rs

Vuk IGNJATOVIĆ   Singidunum University
                                   Faculty of Technical Sciences,
                                   Software and Data Engineering
                                   Danijelova 32
                                   SERBIA
                                   E-mail: vuk.ignjatovic.18@singimail.rs

Horia IUGA   "Horea, Cloșca și Crișan" National College
                                   B-dul 1 Decembrie 1918, No. 11, Alba Iulia
                                   ROMÂNIA
                                   E-mail: iugahoria@gmail.com

Iskren IVANOV   University of Ruse
                                   Department of Informatics and Information Technologies
                                   8, Studentska str., Ruse 7017
                                   BULGARIA
                                   E-mail: iskren_damqnov@abv.bg

Aleksandar IVANOVIC   Singidunum University
                                   Faculty of Technical Sciences
                                   Danijelova 32, 11000 Belgrade
                                   SERBIA

Asaf Ali KHAN   Siena State University
                                   Department of Business and law
                                   Piazza San Francesco, 8, Siena
                                   ITALY
                                   E-mail: asafalikhan8@gmail.com

Mariia KHAN   Siena State University
                                   Department of Information Engineering and Mathematics
                                   Via Roma, 56, Siena
                                   ITALY
                                   E-mail: mariiakhan@yandex.ru

Dominik KLOPFER   University of Applied Sciences Wuerzburg-Schweinfurt
                                   Faculty of Computer Science
                                   Sanderheinrichsleitenweg 20, 97074 Würzburg
                                   GERMANY
                                   E-mail: dominik.klopfer@student.fhws.de

Esther van der KNAAP   Univ. of Georgia
                                The Plant Center
                                111 Riverbend Rd
                                Athens GA 30602
                                USA
                                E-mail: EsthervanderKnaap@uga.edu

Nikolai KÖRBER   University of Applied Sciences Landshut
                                Faculty of Computer Science
                                Am Lurzenhof 1, 84036 Landshut
                                GERMANY
                                E-mail: nikkoerber@gmail.com

Sava LAKIĆEVIĆ   Singidunum University
                                Faculty of Technical Sciences,
                                Software and Data Engineering
                                Danijelova 32
                                SERBIA
                                E-mail: sava.lakicevic.18@singimail.rs

Emanuel Andrei LESTYAN   "Aurel Vlaicu" University of Arad
                                Str. Elena Drăgoi, nr. 2,
                                310330, Arad
                                ROMANIA
                                E-mail: lestyanemanuel@gmail.com

Roland MESZAROȘ   "Aurel Vlaicu" University of Arad"
                                Mathematics and Computer Science
                                Bd. Revoluției 77, 310295 Arad, Romania
                                ROMANIA
                                E-mail: rolandmesaros@gmail.com

Andrej MIĆOVIĆ   Singidunum University
                                Faculty of Informatics and Computing
                                Danijelova 32, 11000 Belgrade
                                SERBIA

Branivoj MILJKOVIC   Singidunum University
                                Faculty of Technical Sciences
                                Danijelova 32, 11000 Belgrade
                                SERBIA
                                E-mail: branivoj.miljkovic@gmail.com

Mihai Eugen MIUȚĂ   Lucian Blaga University of Sibiu
                                Faculty of Science - Informatics
                                Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
                                ROMANIA
                                E-mail: miuta.mihai@gmail.com

Paul Adrian NISTOR

Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: paul.nistor13@gmail.com

Nikola PENEVSKI

Singidunum University
Facutly of Informatics and Computing
32 Danijelova St., Belgrade
SERBIA
E-mail: nikola.penevski.16@singimail.rs

Adrian PINTILIE

Transilvania University of Brașov
Faculty of Electrical Engineering and Computer Science
Bulevardul Eroilor 29, Brașov 500036
ROMANIA
E-mail: pintilie_adrian29@yahoo.com

Oriana PRESACAN

Lucian Blaga University of Sibiu
Department of Computers and Information Technology
Str. Emil Cioran 4 550025, Sibiu
ROMANIA
E-mail: orianapresacan@yahoo.com

Rebeca PUI

Aurel Vlaicu University of Arad
Departament of Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2, Arad
ROMÂNIA
E-mail: nsdproject.vio@gmail.com

Ionel Alexandru PUȚ

Aurel Vlaicu University of Arad
Str. Elena Drăgoi, nr. 2,
310330, Arad
ROMANIA
E-mail: ionelput98@gmail.com

Lyudmila SABADASH

Simon Kuznets Kharkiv National University of Economics
Department of Higher Mathematics, Economic and
Mathematical Methods
Nauki avenue, 9-A, Kharkiv, 61166
UKRAINE
E-mail: malyarets@ukr.net

Victoria SĂRĂTEANU

Aurel Vlaicu University
Departament of Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: victoria.sarateanu1@gmail.com

Eduard-Traian ȘTEFĂNESCU   Lucian Blaga University of Sibiu
Faculty of Science - Informatics
Str. Dr. I. Ratiu, No.5-7, Sibiu, 550012
ROMANIA
E-mail: eduard.traian.stefanescu@gmail.com

Milos TODOROVIC   Singidunum University
Faculty of Technical Sciences
Danijelova 32, 11000 Belgrade
SERBIA

Andrei TOROK   Aurel Vlaicu University
Departament of Mathematics and Computer Science
Str. Elena Drăgoi, Nr.2
ROMANIA
E-mail: andrei_tk@yahoo.com

Ira TUBA   Singidunum University
Faculty of Technical Sciences
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: ira.tuba@gmail.com

Milan III TUBA   Singidunum University
Faculty of Technical Sciences
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: milan.tuba.17@singimail.rs

Una TUBA   Singidunum University
Faculty of Technical Sciences
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: tuba.una@gmail.com

Ana VESIĆ   Singidunum University
Faculty of Technical Sciences
Software and Data Engineering
Danijelova 32
SERBIA
E-mail: ana.vesic.18@singimail.rs

Sofia VISA   College of Wooster
Computer Science Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: svisa@wooster.edu

Yuliya VLASOVA

Volga State University of Service,
Department of Applied Informatics in Economics
4 Gagarin st., Togliatti, Samara region, 445677
RUSSIA
E-mail: jdvlasova@yandex.ru

Angelo WILLIAMS

College of Wooster
Computer Science Dept.
308 E. University St.
Wooster, OH 44691
USA
E-mail: awilliams21@wooster.edu

SPONSORS (in alphabetical order)

AUSY Technologies Romania

Asociaţia BIT

CodexWorks technologies

Fundaţia Academia Ardeleană

Global Solutions for Development

IQuest

Keep Calling

NTT Data

PAN FOOD

Omeron Technologies, Romania

ProIT

ROPARDO

Top Tech

VISMA